OM of: US-09-303-518D-131 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 6:41 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-131
Query length: 1344
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564	+	2281.00	4391.99	3.9e-237	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561	+	2221.00	4275.60	1.2e-230	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562	+	2148.00	4133.98	9.2e-223	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563	+	1631.00	3135.07	5.5e-167	32			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560	+	757.00	1448.34	1.0e-72	15			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439	+	648.00	1223.88	1.1e-60	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318	+	648.00	1223.80	1.1e-60	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271	-	582.50	1113.32	6.1e-54	11			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272	-	545.00	1040.47	7.0e-50	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273	-	532.00	1015.25	1.8e-48	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082	+	450.50	840.30	2.5e-39	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375	+	450.50	840.25	2.5e-39	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553	+	438.00	816.21	5.6e-38	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467	+	161.50	280.30	4.1e-08	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343	+	161.50	280.08	4.2e-08	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	+	145.50	238.96	3.6e-06	10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	+	145.50	233.23	4.8e-06	16			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	+	130.00	203.34	0.0002	15			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817	+	129.00	207.17	0.0002	99			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998	+	127.50	215.98	0.0002	38			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146	+	125.00	209.57	0.0004	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955	+	119.00	194.66	0.0019	57			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935	+	117.00	187.61	0.0036	74			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731	+	115.50	185.46	0.0051	69			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820	+	113.50	174.74	0.0115	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB96063	+	113.00	178.55	0.0104	82			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817	-	112.50	175.16	0.0132	99			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY28654	-	112.00	176.91	0.0131	80			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220	-	109.00	173.91	0.0242	63			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY31745	+	108.50	177.85	0.0217	43			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04954	+	108.50	174.31	0.0257	57			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198	-	108.00	155.54	0.0678	24			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	-	107.50	159.69	0.0602	15			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242	-	106.50	170.45	0.0422	57			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG93052	+	106.50	167.58	0.0484	72			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21954	-	106.50	162.58	0.0614	10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504	+	106.00	161.80	0.0689	10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	+	106.00	160.16	0.0745	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB60656	-	106.00	153.41	0.1026	20			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48589	-	105.50	173.82	0.0420	37			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU39645	-	105.00	174.04	0.0450				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44890	+	105.00	168.42	0.0588				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21970	+	105.00	167.46	0.0616				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26240	+	105.00	159.63	0.0893				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239	+	105.00	158.01	0.0965				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU01253	+	104.50	172.19	0.0532				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23390	-	104.50	157.28	0.1080				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU35592	+	104.00	168.36	0.0689				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69806	-	104.00	151.41	0.1544				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:AAP50073	-	103.50	158.90	0.1170				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:AAR80144	-	103.50	158.90	0.1170				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY68009	-	103.50	158.90	0.1170				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23389	-	103.50	155.49	0.1375				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51653	+	103.00	169.31	0.0771				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146	-	103.00	166.89	0.0865				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY00243	+	103.00	162.62	0.1059				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG28653	+	103.00	157.50	0.1352				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM41160	+	103.00	157.30	0.1365				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22375	+	103.00	150.43	0.1892				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39065	+	103.00	149.49	0.1979				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY00238	+	103.00	147.93	0.2131				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY00240	+	103.00	147.93	0.2131				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY00242	+	103.00	147.93	0.2131				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW34055	+	102.50	165.51	0.0999				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51580	+	102.50	161.53	0.1207				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG04814	-	102.00	158.70	0.1493				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:AAR67371	+	102.00	157.45	0.1584				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80699	+	102.00	157.01	0.1618				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB11821	+	102.00	155.36	0.1750				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM79850	+	102.00	155.36	0.1750				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM78866	+	102.00	154.79	0.1798				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51862	+	101.50	175.06	0.0742				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	-	101.50	147.87	0.2703				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39374	+	101.00	153.50	0.2236				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU58082	+	100.50	166.11	0.1328				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU38243	+	100.50	160.24	0.1755				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44897	+	100.50	159.67	0.1803				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44899	+	100.50	159.67	0.1803				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:AAP60516	+	100.50	159.05	0.1857				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB03127	+	100.50	146.23	0.3417				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB03128	+	100.50	143.91	0.3815				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80217	+	100.00	160.74	0.1853				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80218	+	100.00	159.99	0.1920				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80216	+	100.00	158.67	0.2045				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE08643	+	100.00	153.18	0.2655				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44889	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44891	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44893	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44894	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44895	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44896	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44898	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44902	+	99.50	157.73	0.2313				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY97316	-	99.50	148.88	0.3522				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG80035	+	99.00	157.38	0.2543				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG80043	+	99.00	157.38	0.2543				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44892	+	99.00	156.87	0.2605				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE08645	+	99.00	151.24	0.3405				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB62399	-	99.00	148.96	0.3795				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY08400	+	98.50	157.50	0.2734				

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564

seq_documentation_block:
ID AAY38564 standard; Protein; 447 AA.
XX
AC AAY38564;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigen encoded by ORF22.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

```
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12028.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 125; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality: 2281.00              Length:    447
              Ratio:  5.103                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-131 x AAY38564   ..

 Align seg 1/1  to: AAY38564  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

     751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

     801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

     851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

     901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

     951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

    1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

    1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

    1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

    1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

     1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

     1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

     1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
          |||||||||||||||||||||||||||||||||||||||||
      434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561

seq_documentation_block:
ID   AAY38561 standard; Protein; 447 AA.
XX
AC   AAY38561;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12026.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
           Quality: 2221.00            Length:    447
             Ratio:  4.980              Gaps:      0
 Percent Similarity:  99.776   Percent Identity: 96.197
```

```
alignment_block:
 US-09-303-518D-131 x AAY38561   ..

 Align seg 1/1  to: AAY38561  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||::::::||||||||||||||||||||||||||||||||||||||||
      17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||:::|||||||||||||||||||:::||||||||||:::|||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
         ||||||||||||||:::|||||||||||||||:::||||||:::|||::
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
         :||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
         ||||||||||||||||||||||||||||||||||:::|||||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250

     751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
         ||||||:::|||||||||||||||||||||||||:::|||||||||||||
     251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

     801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
         |   ||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
```

```
    851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
        ||||||||:::|||||||||||||||||||||:::|||||||||||||||
    284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300

    901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
        ||||||||||||||||||||||||:::|||||||||||||||||||||||
    301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

    951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

   1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

   1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
        ||||||||||||||||||||||||||||||||||||||||:::||||||
    351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367

   1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

   1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

   1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

   1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

   1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
        |||||||||||||||||||||||||||||||||||||||||
    434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562

seq_documentation_block:

```
ID   AAY38562 standard; Protein; 447 AA.
XX
AC   AAY38562;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12027.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;
```

alignment_scores:

```
            Quality: 2148.00              Length:    447
              Ratio:  4.927                Gaps:      0
 Percent Similarity:  97.539    Percent Identity: 93.289
```

alignment_block:

US-09-303-518D-131 x AAY38562 ..

Align seg 1/1 to: AAY38562 from: 1 to: 447

```
      1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

     51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
        ||||||||||||||||||||:::|||||||||||||||||||||||||||
     17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34

    101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
        ||||||||:::|||||||||||:::|||||:::|||||||||:::|||
     34 luGluTyrAlaGlyMetArgPro***MetLysValLysGluGlyAspAla 50

    151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
        |||||||||||||||||||||||||||||||||   ||||||||||||
     51 ValLysLysGlyGlnValLeuPheGluAspLysLys***ProGlyValVa 67

    201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
        |||||||||||||:::|||||||||||||||||||||||||||||||||
     67 lPheThrAlaProValSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

    251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

    301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
        ||||||||||||||:::|||||||||||||||:::||||||:::   ::
    101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGly***Gl 117

    351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
        :         ||||||||||||||||||||||||||||||||||||:::|
    117 u*********AsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArg***A 134

    401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

    451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
        ||||||||||||||||||||||||||||||||||||:::|||:::||
    151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValValIl 167

    501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
        ||||||||||:::   |||||||:::|||      ||||||||||||||
```

```
 167 eLysGluAla******AspPheArgArg******LeuValLeuSerArgL 184

 551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||:::|||||||||||||||||||||||||||:::||||||||||||
 251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     |   ||||||||||||||||||||||||||||||||||||||||||||||||
 267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     |||||||:::|||||||||||||||||||||||||||||||||||||||||
 284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGC 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

1051 ACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

1101 CGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

1151 TAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGlu***GlyP 434

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
     |||||||||||||||||||||||||||||||||||||||||
 434 roLeuLeuArgLysValLeuGluThr***GluLysGluGly 447
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563

seq_documentation_block:
ID AAY38563 standard; Protein; 322 AA.
XX
AC AAY38563;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigen encoded by a partial ORF22.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 124-125; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 322 AA;

alignment_scores:
Quality: 1631.00 Length: 322
Ratio: 5.065 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-303-518D-131 x AAY38563 ..

Align seg 1/1 to: AAY38563 from: 1 to: 322

```
   1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

  51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

 201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

 251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

 301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

 351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

 501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

 551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

 601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

 651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

 701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

 751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

 801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

 851 TGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTT 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAAT 966
     ||||||||||||||||
 317 uGlyArgTyrHisAsn 322

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560

seq_documentation_block:
ID   AAY38560 standard; Protein; 158 AA.
XX
AC   AAY38560;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12025.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 122; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   158 AA;

alignment_scores:
            Quality:  757.00              Length:    158
              Ratio:  4.822                 Gaps:      0
 Percent Similarity:  99.367   Percent Identity: 93.671

alignment_block:
 US-09-303-518D-131 x AAY38560   ..

 Align seq 1/1  to: AAY38560  from: 1  to: 158

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||:::::|||||||||||||||||||||||||||||||||||||||||
      17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||:::|||||||||||||||||||:::|||||||||:::|||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84
```

```
      251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
          ||||||||||||||||||||||||||||||||   ||||||||||||
       84 ysArgValLeuGlnSerValValIleAlaValGlu***AsnAspGluIle 100

      301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
          ||||||||||||||||:::||||||||||||||:::||||||:::|||::
      101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

      351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
          :||||||||||||||||||||||||||||||||||||||||||||||||||
      117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

      401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

      451 GTCAATGCGATGGACACCAATCCG 474
          |||||||||||||||||||||||||
      151 ValAsnAlaMetAspThrAsnPro 158

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439

seq_documentation_block:
ID   AAY34439 standard; Protein; 451 AA.
XX
AC   AAY34439;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG1.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91657.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 417-418; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   451 AA;


alignment_scores:
           Quality:  648.00            Length:    452
             Ratio:  2.189               Gaps:      7
 Percent Similarity:  65.487   Percent Identity: 34.071

alignment_block:
 US-09-303-518D-131 x AAY34439   ..

 Align seg 1/1  to: AAY34439  from: 1  to: 451

        1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
          :::||||||   ||||||||||||   |||   :::   |||:::|||
        4 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 20

       51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC...GCGTTGCTTG 97
                :::         ||||||   :::   :::   |||::::::
       20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 37

       98 GCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAA 147
            :::::::::   |||:::   |||::::::      :::   |||:::
       37 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 53

      148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGT 197
          |||:::   |||::::::|||::::::   |||      |||   ::
       54 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 70

      198 AGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCG 247
          :   ||||||:::|||:::||||||::::::   |||::::::||||||
       70 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 87

      248 AAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
            ||||||   :::   |||:::   :::      :::|||   :::|||
       87 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103

      298 ATCGAG...TTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAG 344
          |||   |||           |||:::|||         ||||||::
      104 TyrGluSerProValGlyAspProSerAla.....LeuSerAl 117

      345 CGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTC 394
          :|||:::::::::::::   |||:::::|||||||:::|||      ::::
      117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 134

      395 GCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCC 444
          ::   ||||||::::::   :::   :::   |||      |||
      134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150

      445 ATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGT 494
          |||::::::::::|||      |||   |||||||||   |||
      151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167

      495 CATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGA 544
           |||::::::         :::   ::::::   ::::::   :::|||:
      167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184

      545 GCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGAC 594
          :::::||||||      |||:::::|||            |||::::::
      184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200

      595 GTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCC 644
          :::      :::||||||      :::|||:::|||         |||||
      201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214

      645 GCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCG 694
          ||||||||||||      |||:::   |||:::      :::|||::::
      214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231
```

```
 695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATC 744
     ::   ::::::||||||||||||:::::::      |||:::|||:::||| 
 231 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247

 745 GGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTT 794
     ||||||:::::::::|||||||:::   :::      |||:::::|||:: 
 248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264

 795 GGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTG 844
     :   |||   ::::::               :::|||   ::   |||: 
 264 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 281

 845 CGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTT......GACGCGGAC 888
     :::::|||               |||:::|||:::      :::::: 
 281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297

 889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGC 938
     :::||||||||||:::|||:::||||||:::||| 
 298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 314

 939 GCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
      :::   :::|||:::         :::|||||||:::||||||   |||| 
 314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331

 989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
     ||   :::      ||||||||||||||:::||||||:::   |||::: 
 331 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 347

1036 TACTCCATCACGCGCACCACTCTCGGCCATTTCCTA...AAAAACAAACT 1082
     ||||||:::::::|||:::   ::::::   :::      ||||||||| 
 348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 364

1083 CTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGA 1132
      :::   :::   :::   ::::::||||||:::|||||||||::: 
 364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 381

1133 TCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTT 1182
       :::   |||:::|||||||:::|||:::||||||   |||      ||| 
 381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTG 1232
     |||:::   :::|||:::   |||   |||:::   :::   |||||| 
 398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 414

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
         |||:::      ||||||:::|||   |||:::|||||      : 
 414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431

1283 GCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAG' 1332
     ::|||   |||          :::::::|||:::   |||:::   ::: 
 431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447

1333 AAGGAA 1338
     |||||| 
 448 LysGlu 449
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318

seq_documentation_block:

```
ID   AAY34318 standard; Protein; 454 AA.
XX
AC   AAY34318;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG1.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91536.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 277; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   454 AA;
```

alignment_scores:

```
            Quality:  648.00              Length:    452
              Ratio:  2.189                 Gaps:      7
 Percent Similarity:  65.487   Percent Identity: 34.071
```

alignment_block:

US-09-303-518D-131 x AAY34318 ..

Align seg 1/1 to: AAY34318 from: 1 to: 454

```
   1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
     :::||||||   ||||||||||||   |||   :::   |||:::||| 
   7 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 23

  51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC...GCGTTGCTTG 97
           :::        ||||||   :::   :::   |||:::::: 
  23 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 40

  98 GCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAA 147
       :::::::::   |||:::   |||::::::      :::   |||::: 
  40 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 56

 148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGT 197
       |||:::   |||:::::|||::::::   |||      |||   :: 
  57 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 73

 198 AGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCG 247
     :   ||||||:::|||:::||||||::::::   |||:::::||||||| 
  73 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 90

 248 AAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
```

```
 90 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 106

298 ATCGAG...TTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAG 344
107 TyrGluSerPheProValGlyAspProSerAla.........LeuSerAl 120

345 CGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTC 394
120 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 137

395 GCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCC 444
137 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 153

445 ATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGT 494
154 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 170

495 CATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGA 544
170 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 187

545 GCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGAC 594
187 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 203

595 GTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCC 644
204 LeuGlyLeuHisAsnAlaGluIleValGluValHis.........GlyPr 217

645 GCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCG 694
217 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 234

695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATC 744
234 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 250

745 GGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTT 794
251 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 267

795 GGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTG 844
267 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 284

845 CGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTT......GACGCGGAC 888
284 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 300

889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGC 938
301 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 317

939 GCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
317 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 334

989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
334 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 350

1036 TACTCCATCACGCGCACCACTCTCGGCCATTTCCTA...AAAAACAAACT 1082
351 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 367

1083 CTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGA 1132
367 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 384

1133 TCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTT 1182
```

```
384 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 400

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTG 1232
401 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 417

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
417 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 434

1283 GCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAG 1332
434 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 450

1333 AAGGAA 1338
451 LysGlu 452
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271

seq_documentation_block:

ID AAY75271 standard; Protein; 119 AA.
XX
AC AAY75271;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 628 protein sequence SEQ ID NO:2016.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54033.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1003; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
```
```

```
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   119 AA;


alignment_scores:
            Quality:  582.50              Length:    120
              Ratio:  4.895                Gaps:      1
 Percent Similarity:  99.167   Percent Identity: 99.167

alignment_block:
 US-09-303-518D-131/rev x AAY75271   ..

 Align seg 1/1  to: AAY75271  from: 1  to: 119

     674 ATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGT 625
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTAC 575
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 lSerIleLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

     574 ACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGT 525
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 isThrTrpIleLeuArgSerValArgArgLeuAsnThrAsnArgProArg 50

     524 TTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAG 475
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuMetMetThrValGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTA 425
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT 375
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

     374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC 325
         |||   ||||||||||||||||||||||||||||||||||||||||||||
     101 Asp...IleArgLeuArgArgThrPheSerLeuLeuAsnPheAlaSerAl 116

     324 TTCAGGTACG 315
         ||||||||||
     116 aSerGlyThr 119

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272

seq_documentation_block:
ID   AAY75272 standard; Protein; 120 AA.
XX
AC   AAY75272;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2018.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54034.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics
XX
PS   Claim 2; Page 1003; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;


alignment_scores:
            Quality:  545.00              Length:    119
              Ratio:  4.698                Gaps:      0
 Percent Similarity:  97.479   Percent Identity: 93.277

alignment_block:
 US-09-303-518D-131/rev x AAY75272   ..

 Align seg 1/1  to: AAY75272  from: 1  to: 120

     674 ATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGT 625
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTAC 575
         ||||:::||||||||||||||||||||||||||||||||||||||||||:
      17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuG 34

     574 ACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGT 525
         ::|||||||||||||||||||:::|||||||||||||||||||||||||
      34 lnThrTrpIleLeuArgSerValLysArgLeuAsnThrAsnArgProArg 50

     524 TTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAG 475
         ||||||||||||||||||||||||:::|||||||||||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuIleMetThrValGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTA 425
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT 375
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

  374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC 325
      |||   ||||||||||||||||||   |||   |||:::||||||||||
  101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaSerAl 117

  324 TTCAGGT 318
      |||||||
  117 aSerGly 119

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273

seq_documentation_block:
ID   AAY75273 standard; Protein; 120 AA.
XX
AC   AAY75273;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2020.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Sqarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54035.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1004; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;

alignment_scores:
           Quality:  532.00            Length:    119
             Ratio:  4.626               Gaps:      0
 Percent Similarity:  96.639   Percent Identity: 89.916

alignment_block:
 US-09-303-518D-131/rev x AAY75273   ..

 Align seg 1/1  to: AAY75273  from: 1  to: 120

  674 ATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGT 625
      ||||||||||||||||||||||||||||||||||||||||||||||||||
    1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

  624 TTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTAC 575
      ||||:::||||||||||||||||||||||||||||||||||||||||||
   17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

  574 ACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGT 525
      ||||||||||||||||||||||:::||||||||||||:::::||||||
   34 isThrTrpIleLeuArgSerValLysArgLeuAsnThrSerLysProArg 50

  524 TTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAG 475
      |||||||||||||||||||||||:::   |||:::|||||||||||||
   51 LeuLysSerSerAlaAlaSerLeuIleThrThrThrGlySerAlaAlaSe 67

  474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTA 425
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

  424 CGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT 375
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

  374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC 325
      |||   ||||||||||||||||||   |||   |||:::||||||:::||
  101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaAsnAl 117

  324 TTCAGGT 318
      |||||||
  117 aSerGly 119

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082

seq_documentation_block:
ID   AAY82082 standard; Protein; 467 AA.
XX
AC   AAY82082;
XX
DT   01-JUN-2000  (first entry)
XX
DE   Chlamydia pneumoniae antigen CPN100605 protein SEQ ID NO:2.
XX
KW   Chlamydia pneumoniae; antigen; CPN100605 protein; immunisation;
KW   vaccine; infection; antibacterial; antiinflammatory; bronchitis;
KW   community acquired pneumonia; upper respiratory tract infection;
KW   sinusitis.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200006742-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-IB01331.
XX
PR   27-JUL-1998;   98US-0094195.
PR   26-JUL-1999;   99US-0361443.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
```

PI Murdin AD, Oomen RP;
XX
DR WPI; 2000-205466/18.
DR N-PSDB; AAZ95378.
XX
PT Chlamydia pneumoniae antigens used for immunization and protection
PT against Chlamydia diseases -
XX
PS Claim 6; Fig 1; 48pp; English.
XX
CC The present sequence represents the Chlamydia pneumoniae antigen
CC CPN100605 protein. The CPN100605 protein has antibacterial and
CC antiinflammatory activities. The Chlamydia pneumoniae CPN100605
CC polynucleotide and protein can be used in vaccination methods for
CC preventing and treating Chlamydia infection (e.g. infections caused by
CC C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum). The
CC polynucleotide can be used to produce the protein recombinantly, in the
CC construction of vaccine vectors, as a vaccine agent, and in the
CC construction of an attenuated Chlamydia strain. The protein are also be
CC useful as a vaccine agent, and for the preparation of medicaments for
CC treating or preventing Chlamydia infection, e.g. community acquired
CC pneumonia, and upper respiratory tract infections such as bronchitis and
CC sinusitis.
XX
SQ Sequence 467 AA;

alignment_scores:
Quality: 450.50 Length: 470
Ratio: 1.532 Gaps: 16
Percent Similarity: 62.553 Percent Identity: 29.149

alignment_block:
US-09-303-518D-131 x AAY82082 ..

Align seg 1/1 to: AAY82082 from: 1 to: 467

```
  4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCG...GA 50
    |||   :::::::::|||||||:::|||   :::   |||   |||   ||
  3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 19

 51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
    |:::   :::|||:::                    :::
 19 uSerGlyPheTyrAsn.....................LysIleAspP 28

101 AAGAATATGTCGGC......ATGCGCCCC.................TCG 126
      |||:::|||:::      :::|||||                  :::
 28 roGluPheValSerIleAspLeuArgProPheGlnProLeuSerLeuLys 44

127 ATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGA 176
    :::|||:::::::::|||:::|||||   :::|||      :::   ||
 45 LeuLysValGluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGl 61

177 AGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAA 226
    |   |||      |||::::::   :::|||:::   :::||||||   :
 61 uTyrLysHisProAsnThrTyrIleThrSerHisValSerGlyValV 78

227 TCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATT 276
    :::::||||||:::||||||:::|||||   |||   :::|||:::|||
 78 alThrAlaIleArgArgGlyAsnLysArgSerLeuLeuAspValIleIle 94

277 ...GCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
          :::   |||      :::   |||:::      |||      ::
 95 LysLysThrProGlyProThrSerThrGluTyr...ThrTyrAspLeuGl 110

324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAAT 373
    ::::|||::::::   :::   :::::::::   :::      ::::
110 nThrLeuSerArgSerAspLeuSerGluIlePheLys.....GluA 124

374 CAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCC 423
    ::|||||:::::   ::::::   ||||||||||:::   |||||||||
124 snGlyLeuPheAlaLeuIleLysGlnArgProPheAsp...IleProAla 139

424 GTAGATGCCGAG...CCGTTCGCCATCTTCGTCAATGCGATGGACACCAA 470
    :::   ::::::   |||      :::|||:::|||      |||:::::
140 IleProThrGlnThrProArgAspValPheIleAsnLeuAlaAspAsnAr 156

471 TCCGCTGGCTGCCGACCCTACGGTCATCATC.................A 502
    :|||::::::   :::|||         :::                 :
156 gProPheThrProSerProGluLysHisLeuAlaLeuPheSerSerArgG 173

503 AAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTG 552
    ::|||:::      |||      |||:::   ::::::::::::|||
173 luGluGlyPheTyrValPheValValGlyValArgAlaIleAlaLysLeu 189

553 ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTC 602
              :::   |||:::                    :::|||::
190 PheGlyLeuArgProHisIleValPheArgAspArgLeuThrLeuProTh 206

603 TGAAAATGCTGCCAATATC...GAAACACATGAATTTGGCGGCCCGCATC 649
    :::::::      :::|||   :::   |||      ::|||||||   |
206 rGlnGluLeuLysThrIleAlaHisLeuHisThrValSerGlyProPheP 223

650 CTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCG 699
    ||:::|||      :::   |||||||||   :::   |||:::
223 roSerGlySerProSerIleHisIleHisSerValAlaProIleThrAsn 239

700 AATAAA...ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGG 746
    :::|||   :::|||:::|||:::::::::|||||||||:::::|||||
240 GluLysGluValValPheThrLeuSerPheGlnAspValLeuThrIleGl 256

747 ACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGG 796
    |:::|||||||:::   ||||||:::      |||:::|||:::||||||:
256 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 273

797 GCGGC......CTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840
    ::|||      |||:::   :::      |||   :::   |||:::
273 laGlyThrAlaLeuLysSerSerLeuArgArgTyrValIleThrThrLys 289

841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAA 890
    ||||||:::   |||:::|||         :::   ::::::   :::::
290 GlyAlaSerPheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAs 306

891 CCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGC 940
    :   :::||||||||||:::   |||:::|||   :::::::::      :
306 pThrLeuIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluG 323

941 ATGAT...TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAA 987
    :::::   :::|||||         :::::|||||||||:::::::::
323 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 339

988 GGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA 1037
          ::::::||||||||:::::::::              :::|||
340 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 356

1038 CTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCA 1087
     :::::::|||:::|||   |||:::   |||:::|||:::|||
356 oThrPheThrLysThrTyrLeuSerGlyPhePheLysLysLysArg...T 372

1088 AGTTCACG.........ACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
       :::|||         |||   ::::::|||      |||   ::::::
372 hrTyrThrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIle 388

1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
                |||:::::::|||||||||:::   |||      :::
389 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 405

1179 GCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGG 1228
     |||::::::   :::|||:::   :::   |||   ||||::   ||||
405 oLeuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 422

1229 GTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGC 1278
     ||   |||||||::      |||||||:::||||||   :::::::::
422 lyPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleAsp 438

1279 CCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCAT 1328
```

```
      |||:::|||   |||        ::::::::::::   |||
  439 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGluTy 455

 1329 TGAGAAGGAA 1338
         ||||||
  455 rAlaLysGlu 458
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375

seq_documentation_block:

ID AAY35375 standard; Protein; 469 AA.
XX
AC AAY35375;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1170-1171; Disclosure; 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotides sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 469 AA;

alignment_scores:

```
            Quality:  450.50              Length:    470
              Ratio:  1.532                 Gaps:     16
 Percent Similarity:  62.553   Percent Identity: 29.149
```

alignment_block:

US-09-303-518D-131 x AAY35375 ..

Align seg 1/1 to: AAY35375 from: 1 to: 469

```
      4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCG...GA 50
        |||   :::::::::||||||:::|||   :::   |||   |||   ||
      5 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 21

     51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100

    |:::   :::|||:::                              :::
 21 uSerGlyPheTyrAsn.....................LysIleAspP 30

101 AAGAATATGTCGGC......ATGCGCCCC.................TCG 126
      |||:::|||:::      :::||||||                  :::
 30 roGluPheValSerIleAspLeuArgProPheGlnProLeuSerLeuLys 46

127 ATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGA 176
    :::|||:::::::::|||:::||||||   :::|||      :::   ||
 47 LeuLysValGluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGl 63

177 AGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAA 226
    |   |||      |||::::::   :::|||:::   :::||||||   :
 63 uTyrLysHisPheProAsnThrTyrIleThrSerHisValSerGlyValV 80

227 TCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATT 276
    :::::||||||:::||||||:::||||||   |||   :::|||:::|||
 80 alThrAlaIleArgArgGlyAsnLysArgSerLeuLeuAspValIleIle 96

277 ...GCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGA 323
       :::   |||      :::   |||:::      |||      ::
 97 LysLysThrProGlyProThrSerThrGluTyr...ThrTyrAspLeuGl 112

324 AGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAAT 373
    ::::|||::::::   :::   :::::::::   :::         ::::
112 nThrLeuSerArgSerAspLeuSerGluIlePheLys.....GluA 126

374 CAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCC 423
    ::||||||::::::   ::::::   |||||||||:::   |||||||||
126 snGlyLeuPheAlaLeuIleLysGlnArgProPheAsp...IleProAla 141

424 GTAGATGCCGAG...CCGTTCGCCATCTTCGTCAATGCGATGGACACCAA 470
    :::   ::::::   |||      :::|||:::|||      |||:::::
142 IleProThrGlnThrProArgAspValPheIleAsnLeuAlaAspAsnAr 158

471 TCCGCTGGCTGCCGACCCTACGGTCATCATC.................A 502
    :|||::::::   :::|||         :::                 :
158 gProPheThrProSerProGluLysHisLeuAlaLeuPheSerSerArgG 175

503 AAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTG 552
    ::|||:::         |||      |||:::   ::::::::::::|||
175 luGluGlyPheTyrValPheValValGlyValArgAlaIleAlaLysLeu 191

553 ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTC 602
          :::   |||:::                          :::||||:
192 PheGlyLeuArgProHisIleValPheArgAspArgLeuThrLeuProTh 208

603 TGAAAATGCTGCCAATATC...GAAACACATGAATTTGGCGGCCCGCATC 649
    :::::::      :::|||   :::   |||      :::||||||   |
208 rGlnGluLeuLysThrIleAlaHisLeuHisThrValSerGlyProPheP 225

650 CTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCG 699
    ||:::|||      :::   |||||||||   :::   |||:::
225 roSerGlySerProSerIleHisIleHisSerValAlaProIleThrAsn 241

700 AATAAA...ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGG 746
    :::|||   :::|||:::|||:::::::::|||||||||:::::|||||
242 GluLysGluValValPheThrLeuSerPheGlnAspValLeuThrIleGl 258

747 ACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGG 796
    |:::|||||:::   ||||||:::      |||:::|||:::||||||:
258 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 275

797 GCGGC......CTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840
    ::|||      |||:::   :::      |||   :::   |||:::
275 laGlyThrAlaLeuLysSerSerLeuArgArgTyrValIleThrThrLys 291

841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAA 890
    ||||||:::   |||:::|||         :::   ::::::   :::::
292 GlyAlaSerPheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAs 308

891 CCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGC 940
    :   :::||||||||||:::   |||:::|||   :::::::::      :
```

```
    308 pThrLeuIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluG 325

    941 ATGAT...TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAA 987
        :::::   :::||||||         :::::::||||||||:::::::::
    325 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 341

    988 GGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA 1037
            ::::::||||||||:::::::::             :::|||
    342 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 358

   1038 CTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCA 1087
         :::::|||:::|||   |||:::   |||:::|||:::|||
    358 oThrPheThrLysThrTyrLeuSerGlyPhePheLysLysLysArg...T 374

   1088 AGTTCACG.........ACAGCCGTCAACGGCGGCGACCGCGCCATGGTA 1128
          :::|||         |||   ::::::|||      |||   ::::::
    374 hrTyrThrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIle 390

   1129 CCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT 1178
                  |||::::::|||||||||:::   |||      :::
    391 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 407

   1179 GCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGG 1228
         |||::::::   :::|||:::   :::   |||   |||:::   ||||
    407 oLeuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 424

   1229 GTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGC 1278
        ||   ||||||:::      ||||||:::||||||   :::::::::
    424 lyPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleAsp 440

   1279 CCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCAT 1328
        |||:::|||   |||         ::::::::::   |||
    441 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGluTy 457

   1329 TGAGAAGGAA 1338
          ||||||
    457 rAlaLysGlu 460
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553

seq_documentation_block:

```
ID   AAY37553 standard; Protein; 461 AA.
XX
AC   AAY37553;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Chlamydia trachomatis surface exposed protein.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 1216; 1755pp; English.
XX
CC   AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC   of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC   vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the invention
CC   may be of use in treating these diseases.
XX
SQ   Sequence   461 AA;
```

alignment_scores:

```
            Quality:  438.00              Length:    463
              Ratio:  1.490                 Gaps:     14
 Percent Similarity:  63.499   Percent Identity: 28.078
```

alignment_block:

```
 US-09-303-518D-131 x AAY37553   ..

 Align seq 1/1  to: AAY37553  from: 1  to: 461

       4 ATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCA 53
         |||   :::::::::||||||:::|||   :::   |||   |||:::::
       3 IleValValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGl 19

      54 AGTCATTTATGACGGC.........CCGGCCATTACCGAAGTCGCGTTGC 94
         :     :::   |||         |||:::   :::::|||   |||
      19 uSerGlyPheCysGlyLysValAspProThrTyrValSerValAspLeu. 35

      95 TTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGT 144
            :::   ::::::   :::      ::::::|||:::      |||
      36 .....ArgProPheAlaProLeuProLeuGluValLysValThrProGly 50

     145 GAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
         :::   |||      |||:::   |||   |||   |||         ||
      51 AspGlnValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerGl 67

     195 CGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTG 244
         ||||   :::|||:::|||:::::|||:::::::::   |||:::||||
      67 yValPheIleThrSerProValAspGlyGluValValGluIleArgArgG 84

     245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA...GGCAAC 291
         ||:::||||||:::|||   ::::::||||||     :::   |||
      84 lyAsnLysArgAlaLeuLeuGluIleValIleLysLysLysProGlyIle 100

     292 GACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAG 341
         ::::::   :::|||:::                |||   :::|||::
     101 SerGlnThrLysPheSer.............TyrAspLeuGlnSerLeuTh 113

     342 CAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGC 391
         :::::::   :::   :::   :::   ::::::||||||::::::   :
     113 rGlnLysAspLeuLeuGluValPheLysLysGluGlyLeuPheAlaLeuP 130

     392 TTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAG...CCG 438
         :::::   |||||||||:::   |||||||||:::   :::::   |||
     130 heLysGlnArgProPheAsp...IleProAlaLeuProThrGlnSerPro 145

     439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG......GCTGC 482
               :::|||:::|||      |||:::::::|||:::      :::::
     146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerVa 162

     483 CGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC.............T 520
         ::::      :::   :::   ::::::   ||||||            |
     162 lGluLysHisLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIleP 179

     521 TCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
```

```
     ||      |||:::   ::::::::::::|||         |||   |||
  179 heValValGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195

  571 GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATAT 620
      :::   ::::::            :::|||:::::::::   ::::::||
  196 IleIleSerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIl 212

  621 C...GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCA 667
      |   :::   |||   :::   ||||||   |||:::|||      :::|
  212 eAlaHisLeuHisThrIleAspGlyProPheProSerGlySerProSerT 229

  668 CGCACATTCATTTCATCGAGCCAGTC...GGCGCGAATAAAACCGTGTGG 714
      ||||||||||||   |||       :::   :::   :::   :::|||:::
  229 hrHisIleHisHisIleAlaArgIleArgAsnGluArgAspValValPhe 245

  715 ACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGG 764
      ||||||:::::|||:::|||:::::||||||:::||||||:::   ||
  246 ThrIleSerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGl 262

  765 CCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGC......CTGCAAG 808
      |   :::       :::::::::|||||||||:::|||.      |||
  262 yPheValLeuGlyGlnGlnIleValAlaLeuAlaGlySerAlaLeuProP 279

  809 TCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAA 858
        ::::::   :::   |||   |||:::   |||||||:::   |||:::
  279 roSerGlnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheSerAsp 295

  859 CTTACCGCCGGCGAATTGGTTGACGCGGACAAC...CGCGTGATTTCCGG 905
      |||           ::::::   ::::::|||:::      :::|||||||
  296 LeuLeuProLysAspIlePheSerSerAspGluIleThrLeuIleSerGl 312

  906 TTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGAT...TATTTGG 952
      |:::   |||:::|||   :::::::::       ::::::      ||||
  312 yAspProLeuThrGlyArgLeuCysLysLysGluGluAsnProCysLeuG 329

  953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
      ||          :::   |||:::::::::   :::   :::::::::|||
  329 lyMetArgAspHisThrIleThrLeuLeuProAsnProLysThrArgGlu 345

 1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCAC 1052
         |||::::::::::            :::|||   ::::::||||||||
  346 SerPheSerPheLeuArgLeuGlyTrpAsnLysLeuThrValThrArgTh 362

 1053 CACTCTCGGCCATTTCCTAAAAAACAAA......CTCTTCAAGTTCACGA 1096
      |   |||:::   |||:::|||:::|||      ::::::   :::   |
  362 rTyrLeuSerGlyPhePheLysArgLysArgValPheMetAspMetAspT 379

 1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
      ||   ::::::|||      |||   ::::::            ||||||
  379 hrAsnMetHisGlyGluLysArgProIleIleAspAlaGluIleTyrGlu 395

 1147 CGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAAT 1196
      ||||||      :::   :::       :::|||:::::::::   |||
  396 ArgValSerAlaIleProValProValAlaLeuIleIleLysAlaLeuGl 412

 1197 CGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACG 1246
       :::   :::   ::::::|||      ||||||   |||||||:::
  412 uThrGlnAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValAlaP 429

 1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
        ||||||:::||||||   :::|||:::   |||:::|||   |||
  429 roGluAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445

 1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAG 1335
            ::::::::::::   |||         :::|||
  446 PheSerIleValLysGluSerLeuLeuArgThrGlnLys 458

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467

seq_documentation_block:
ID   AAY34467 standard; Protein; 443 AA.
XX
AC   AAY34467;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG122.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91685.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 445-446; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   443 AA;


alignment_scores:
            Quality:  161.50              Length:    490
              Ratio:  0.699                 Gaps:     22
 Percent Similarity:  47.143   Percent Identity: 20.612

alignment_block:
 US-09-303-518D-131 x AAY34467   ..

 Align seg 1/1  to: AAY34467  from: 1  to: 443

      37 GCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGT 86
         |||||||:::|||   :::|||:::          |||      ::::::||
      19 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 32

      87 CGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCA 136
         |:::::          :::::::::|||         :::      :::|
      32 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 49

     137 AGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
         ||:::|||:::   ||||||   |||   ::::::
```

```
 49 ysLysGlyAspGluValLysValGlyThrIleIle..... 60

187 AATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCT.. 234
                    |||   |||:::|||   :::::::|||
 61 .....AlaGlnAlaGlyGlyPheValSerAlaAs 70

235 .ATTCACCGTGGC...GAAAAGCGCGTACTTCAG..... 264
    ||||||   :::   :::   :::|||||||:::
 70 nIleHisSerSerValSerGlyLysValLeuLysIleAspAsnValTyrA 87

265 .....TCAGTCGTGATTGCCGTTGAAGGCAAC 291
                    :::|||   |||:::|||||||||:::
 87 spSerSerGlyTyrProLysProAlaValPheIleSerValGluGlyAsp 103

292 GACGAAATCGAGTTCGAACGCTACGTA..... 318
          |||:::|||:::   :::
104 .....GluTrpGluGluGlyIleAspArgSerProAlaIleValLy 117

319 .....CCTGAAGCGCTGGCAAAATTGAGCAGCGAAA 349
                    |||   :::|||||||:::|||:::
117 sGluCysAsnLeuAspAlaLysGluIleValAlaLysIleSerAlaAlaG 134

350 AAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACC 399
    :::                 ||||||      |||      |||
134 lyIle.....ValGlyLeuGlyGlyAla.....Thr 142

400 CGTCCG.....TTCAGCAAAATCCCTGCCGTAGATGCCGAGCC 437
      |||           :::|||      |||:::      ||||||
143 PheProThrHisValLysLeuSerProProProGlyAsnLysAlaGlu.. 158

438 GTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACC 487
        :::::::::||||||:::::   :::|||      ::::::
159 ....IleLeuIleIleAsnAlaValGluCysGluProTyrLeuThrSerA 174

488 CTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTG 537
       |||::::::   |||   :::|||:::::      |||:::
174 spHisValLeuMetLeuGluHisGlyGluGluIleMetIleGlyValSer 190

538 GTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCA.. 585
    :::|||                 :::   |||:::|||   ||||||:::
191 IleLeuMet.....LysAlaIleGlnValAsnLysAlaValIl 203

586 .GGCGCAGACGTGCCGTCTGAAAATGCT..... 612
     |||::::::      :::::::::|||
203 eGlyValGluAsnAsnLysLysAspAlaIleAlaHisLeuThrLysLeuA 220

613 .....GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCT 651
                  :::|||||||:::      :::         :::|||
220 laThrAlaTyrProGlyIleGluValMetProLeuLysValGlnTyrPro 236

652 GCCGGCTTGAGTGGCACGCACATTCATTTCATC..... 684
    |||   :::         |||      :::
237 GlnGlyGlyGluLysGlnLeuIleAspAlaValIleArgLysGlnValLy 253

685 .....GAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATT 724
                 |||:::::::::::::   :::|||         |||
253 sSerGlyAlaLeuProIleSerThrGlyAlaValValGln.....AsnV 268

725 ATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAAT 774
          |||:::|||:::   :::            ::::::
268 alGlyThrValPheAlaValTyrGluAlaValGlnLysAsnLysProLeu 284

775 ACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCT 824
    :::||||||:::|||:::::   |||   :::::::::::|||
285 ValGluArgIleValThrValThrGlyLysLysLeuSerArgProSerAs 301

825 CTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT.....ACCGCCG 868
     |||   :::   :::|||:::   ::::::   |||      :::||||
301 nLeuLeuValArgIleGlyThrProIleAlaAlaLeuIleGluAlaAlaG 318

869 GCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAAC 918
    ||   |||   :::      :::::::::|||:::|||:::
318 lyGlyLeuProGluAsnThrGlyLysIleIleGlyGlyGly..... 331

919 GGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCA 968

331 ..... 331

969 GATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTG 1018
                                      ::::::|||   :::
332 .....ProMetMetGlyArgAlaL 338

1019 CGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTC 1068
        :::||||||         :::|||:::   :::   |||   :::
 338 euLeuSerProAsp...ValProValThrLysGlySerSerGlyValLeu 353

1069 CTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCG 1118
     :::   ::::::                  ||||||:::        ||
 354 IleLeuAspArg.....GluGluAlaValArgLysProMetAr 366

1119 CGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCC 1168
     |      :::      :::              |||   |||:::   :::
 366 gAspCysIleArgCysAlaLysCysValGlyValCysProMetGlyLeuA 383

1169 TGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCG 1218
       |||:::::::|||:::||||||   :::      :::   ::::::|||
 383 snProAlaPheLeuMetArgAspThrLeuTyrLysSerTrpGluThrAla 399

1219 CAG.....GCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGC 1259
     :::         ::::::   |||:::|||      :::
 400 GluLysGlyAsnValValAspCysIleGluCysGlySer..... 412

1260 TTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGC 1309
        |||||||||:::||||||::::::                  :::|
 413 ....CysSerPheThrCysProAlaAsnArgProLeuLeuAspTyrIleA 428

1310 GCAAAGTGCTGGAAACCATT 1329
     ||::::::   :::|||:::
 428 rgGlnAlaLysLysThrVal 434
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343

seq_documentation_block:

```
ID   AAY34343 standard; Protein; 451 AA.
XX
AC   AAY34343;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG122.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
```

PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR N-PSDB; AAX91561.
XX
PT Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 303; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 451 AA;

alignment_scores:
Quality: 161.50 Length: 490
Ratio: 0.699 Gaps: 22
Percent Similarity: 47.143 Percent Identity: 20.612

alignment_block:
US-09-303-518D-131 x AAY34343 ..

Align seg 1/1 to: AAY34343 from: 1 to: 451

```
 37 GCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGT 86
    ||||||:::|||   :::|||:::         |||      ::::::||
 27 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 40

 87 CGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCA 136
    |::::::         ::::::::::|||         :::      :::|
 40 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 57

137 AGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
    ||:::|||:::   ||||||   |||   ::::::
 57 ysLysGlyAspGluValLysValGlyThrIleIle.............;. 68

187 AATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCT.. 234
                    |||   |||:::|||   :::::::|||
 69 ....................AlaGlnAlaGlyGlyPheValSerAlaAs 78

235 .ATTCACCGTGGC...GAAAAGCGCGTACTTCAG.............. 264
     ||||||   :::   :::   :::|||||||:::
 78 nIleHisSerSerValSerGlyLysValLeuLysIleAspAsnValTyrA 95

265 ......................TCAGTCGTGATTGCCGTTGAAGGCAAC 291
                          :::|||   |||:::||||||||||:::
 95 spSerSerGlyTyrProLysProAlaValPheIleSerValGluGlyAsp 111

292 GACGAAATCGAGTTCGAACGCTACGTA..................... 318
          |||:::|||:::   :::
112 .........GluTrpGluGluGlyIleAspArgSerProAlaIleValLy 125

319 ..................CCTGAAGCGCTGGCAAAATTGAGCAGCGAAA 349
                       |||   :::||||||:::|||:::
125 sGluCysAsnLeuAspAlaLysGluIleValAlaLysIleSerAlaAlaG 142

350 AAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACC 399
      :::                |||||      |||      |||
142 lyIle.................ValGlyLeuGlyGlyAla......Thr 150

400 CGTCCG............TTCAGCAAAATCCCTGCCGTAGATGCCGAGCC 437
     |||             :::|||      |||:::      ||||||
151 PheProThrHisValLysLeuSerProProProGlyAsnLysAlaGlu.. 166

438 GTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACC 487
          :::::::::||||||:::::   :::|||      ::::::
167 ....IleLeuIleIleAsnAlaValGluCysGluProTyrLeuThrSerA 182

488 CTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTG 537
      |||::::::   |||   :::|||:::::      |||:::
182 spHisValLeuMetLeuGluHisGlyGluGluIleMetIleGlyValSer 198

538 GTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCA.. 585
    :::|||              :::   |||:::|||   |||||||:::
199 IleLeuMet............LysAlaIleGlnValAsnLysAlaValIl 211

586 .GGCGCAGACGTGCCGTCTGAAAATGCT..................... 612
     |||::::::      ::::::::::|||
211 eGlyValGluAsnAsnLysLysAspAlaIleAlaHisLeuThrLysLeuA 228

613 ...........GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCT 651
                ::::||||||:::      :::          :::|||
228 laThrAlaTyrProGlyIleGluValMetProLeuLysValGlnTyrPro 244

652 GCCGGCTTGAGTGGCACGCACATTCATTTCATC................. 684
    |||   :::         |||      :::
245 GlnGlyGlyGluLysGlnLeuIleAspAlaValIleArgLysGlnValLy 261

685 ..........GAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATT 724
              |||:::::::::::::   :::|||         |||
261 sSerGlyAlaLeuProIleSerThrGlyAlaValValGln......AsnV 276

725 ATCAAGACGTCATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAAT 774
          |||:::|||:::   :::             ::::::
276 alGlyThrValPheAlaValTyrGluAlaValGlnLysAsnLysProLeu 292

775 ACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCT 824
    :::||||||:::|||:::::   |||   :::::::::::|||
293 ValGluArgIleValThrValThrGlyLysLysLeuSerArgProSerAs 309

825 CTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT......ACCGCCG 868
     |||   :::   :::|||:::   ::::::   |||      :::||||
309 nLeuLeuValArgIleGlyThrProIleAlaAlaLeuIleGluAlaAlaG 326

869 GCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAAC 918
    ||   |||   :::      ::::::::::|||:::|||:::
326 lyGlyLeuProGluAsnThrGlyLysIleIleGlyGlyGly......... 339

919 GGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCA 968

339 ................................................. 339

969 GATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTG 1018
                                        :::::::|||   :::
340 ..............................ProMetMetGlyArgAlaL 346

1019 CGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTC 1068
       :::||||||        :::|||:::   :::   |||   :::
346 euLeuSerProAsp...ValProValThrLysGlySerSerGlyValLeu 361

1069 CTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCG 1118
     :::   ::::::                  ||||||:::         ||
362 IleLeuAspArg............GluGluAlaValArgLysProMetAr 374

1119 CGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCC 1168
     |       :::      :::            |||   |||:::   :::
374 gAspCysIleArgCysAlaLysCysValGlyValCysProMetGlyLeuA 391

1169 TGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCG 1218
      |||::::::|||:::||||||   :::      :::   ::::::|||
391 snProAlaPheLeuMetArgAspThrLeuTyrLysSerTrpGluThrAla 407

1219 CAG.........GCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGC 1259
     :::         ::::::   |||:::|||      :::
408 GluLysGlyAsnValValAspCysIleGluCysGlySer........... 420

1260 TTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGC 1309
```

```
          |||||||||:::||||||::::::                  :::|
    421 ....CysSerPheThrCysProAlaAsnArgProLeuLeuAspTyrIleA 436

   1310 GCAAAGTGCTGGAAACCATT 1329
        ||:::::   :::|||:::
    436 rgGlnAlaLysLysThrVal 442

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813

seq_documentation_block:
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;


alignment_scores:
            Quality:  145.50            Length:    497
              Ratio:  0.677              Gaps:     27
 Percent Similarity:  43.260   Percent Identity: 24.748

alignment_block:
 US-09-303-518D-131 x AAB59813   ..

 Align seg 1/1  to: AAB59813  from: 1  to: 1017

      38 CGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC 87
         |||||||:::   |||   :::   ::::::   :::|||   |||
     101 ArgAlaSerSerSerProLysSerIleSerProLysProArgProThrCy 117

      88 GCGT.........TGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGAT 128
```

```
         |||                |||:::::::::|||:::   :::|||||||
117 sArgProSerProGlyThrAlaArgArgValSerThrThrSerProArg. 133

129 GAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAG 178
    ::::::   ||||||      |||   :::   |||:::
134 SerThrThrGlyArg......ArgTrpSerSerProAla........... 144

179 ACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATC 228
     ::::::::::::|||||:::      ::::::      |||:::::::|
145 .ArgArgSerAlaGlyArgAlaGlyArgAlaGlyCysAlaArgSerSerA 161

229 GCCGCTATTCA...CCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGAT 275
    ||:::   |||   |||   |||:::||||||   |||   |||::::::
161 rgLysThrSerArgProIleArgSerAlaArgProSerCysSerLysSer 177

276 TGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAG 325
        ::::::   :::                                |||   :
178 .ProThrSerValSer.....................AlaPheProProS 187

326 CGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCA 375
    ::   |||:::   |||       :::   |||||||||   :::   |||
187 erProAlaArgAlaSerArgThrArgCysArgArgAsnSerLeuProSer 203

376 GGCTTATGGACTGCGCTTCGCACCCGT.........CCGTTCAGCAAAAT 416
    ::::::       :::       ||||||          |||       |||
204 SerValThrArgSerSerAlaThrArgAlaAlaThrProArgArgLysTh 220

417 CCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCA............ 454
     |||:::          :::       |||   ||||||
220 rProCysCysGlyArg.ThrThrArgProProSerSerThrArgAsnSer 236

455 ..............ATGCGATGGACACCAATCCGCTGG.........CT 480
                  |||||||||:::      ||||||          ::
237 SerArgAlaThrTrpMetArgTrpAsnSerSerArgTrpAsnValArgPh 253

481 GCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGG 530
    :|||::::::      :::|||:::   |||         |||:::   :
253 eProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSerArgG 270

531 CCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAG 580
    ::            ::::::   |||         |||   :::
270 lyArgThrIleCysSerSerSerPro.........SerAlaAlaProThr 283

581 CAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACAT 630
       :::|||:::|||                        |||   |||::::::
284 ProArgAlaArgThrProAlaThrThrProThrProSerSerArgGlnPr 300

631 GAATTTGGCGGCCCGC.....................ATCCTGCCGGCTT 659
     :::   :::||||||                          :::|||
300 oSerGlySerAlaArgProSerProProSerSerSerAlaIleProArgA 317

660 GAGTGGCACGCACA............TTCATTTCATCGAGC......... 688
      :::||||||              |||   |||:::|||
317 rgThrAlaArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAsp 333

689 .........................CAGTCGGCGCGAATAAAACCGTG 711
                             :::|||||||||   :::
334 SerAlaIleArgArgSerSerThrThrArgSerAlaArgSerArgArgAs 350

712 TGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAAC 761
       ||||||     :::|||
350 nThrProSerSerAlaSerThr............................ 357

762 AGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCA 811
                        |||      |||      ::   |||
358 ...................AlaThrAlaPro...ProThrArgLysPro 366

812 ACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT 861
    |||:::   ::::::|||                        |||
367 ThrThrGlySerThrCys.....................CysAlaCysAr 376

862 ACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGT 911
     |||   ::::::      :::|||               |||||||||
```

```
 376 gProAlaSerThrValAlaAlaArgArgLysLys......ProValArg. 390

 912 ATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGG..... 952
             |||      ::::::::::|||        |||
 391 .....LysValAlaAlaGlnSerSerArgProSerCysTrpLysSerArg 405

 953 ......GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGC 996
           |||||||||   |||   |||      ::::::|||
 406 SerMetThrAlaThrThrGlyArgThrProThrCysAsnSerAlaArgAr 422

 997 AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCAC 1046
               |||         |||||||||             |||||||
 422 gProValIleSer.........ArgArgSer............ProSerA 432

1047 GC......GCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGT 1090
       ||      :::   |||||||||   |||   :::   :::|||:::|||
 432 rgMetPheGlyArgLeuSerAlaSerSerIleAsnMetArgSerThrSer 448

1091 TCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACT 1140
           |||   |||:::                    |||::::::
 449 ValSerAlaProArgThrCys..................ArgAlaThrSe 459

1141 TATGAGCGCGTAATGCCGTTGGACATCCTGC................... 1171
         ||||||   ||||||      ||||||
 459 rSerSerAlaSerCysArgCysLeuSerCysProGlnSerThrThrAlaA 476

1172 .................CTACCTTGCTTTTGCGCGATTTAATCGTCGGCG 1204
                        |||:::   |||                |||
 476 laTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAla 492

1205 ATACCGACAGCGCGCAGG.....................CTTTGGGTTGC 1233
           |||      |||                  |||   ::
 493 GlyThrThrArgSerArgArgSerSerArgArgThrProSerTrpProSe 509

1234 TTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGG 1283
       :   ||||||   :::::                :::|||::::::
 509 rArgAsnTrpTyrSerArgArgArgAsnThrProSerSerAsnSerAlaL 526

1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGC 1318
       :::   :::|||::::::         ||||||
 526 ysArgArgArgThrGlyLysValSerArgLysCys 537
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:

```
ID   AAB59826 standard; Protein; 1615 AA.
XX
AC   AAB59826;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #3 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1615 AA;
```

alignment_scores:

```
            Quality:  145.50              Length:    497
              Ratio:  0.677                 Gaps:     27
 Percent Similarity:  43.260   Percent Identity: 24.748
```

alignment_block:

US-09-303-518D-131 x AAB59826 ..

Align seg 1/1 to: AAB59826 from: 1 to: 1615

```
  38 CGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC 87
     ||||||:::   |||   :::   ::::::   :::|||   |||
 699 ArgAlaSerSerSerProLysSerIleSerProLysProArgProThrCy 715

  88 GCGT.........TGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGAT 128
       |||                |||:::::::::|||:::   :::|||||||
 715 sArgProSerProGlyThrAlaArgArgValSerThrThrSerProArg. 731

 129 GAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAG 178
     ::::::   ||||||      |||   :::   |||:::
 732 SerThrThrGlyArg......ArgTrpSerSerProAla........... 742

 179 ACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATC 228
      ::::::::::|||||:::      ::::::      |||::::::|
 743 .ArgArgSerAlaGlyArgAlaGlyArgAlaGlyCysAlaArgSerSerA 759

 229 GCCGCTATTCA...CCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGAT 275
     ||:::   |||   |||   |||:::||||||   |||   |||::::::
 759 rgLysThrSerArgProIleArgSerAlaArgProSerCysSerLysSer 775

 276 TGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAG 325
         ::::::   :::                              |||   :
 776 .ProThrSerValSer.....................AlaPheProProS 785

 326 CGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCA 375
     ::   |||:::   |||      :::   |||||||||   :::   |||
 785 erProAlaArgAlaSerArgThrArgCysArgArgAsnSerLeuProSer 801

 376 GGCTTATGGACTGCGCTTCGCACCCGT.........CCGTTCAGCAAAAT 416
     ::::::      :::      ||||||         |||      |||
 802 SerValThrArgSerSerAlaThrArgAlaAlaThrProArgArgLysTh 818

 417 CCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCA............ 454
      |||:::          :::      |||   ||||||
 818 rProCysCysGlyArg.ThrThrArgProProSerSerThrArgAsnSer 834

 455 ..............ATGCGATGGACACCAATCCGCTGG.........CT 480
                    |||||||||:::      ||||||         ::
 835 SerArgAlaThrTrpMetArgTrpAsnSerSerArgTrpAsnValArgPh 851

 481 GCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGG 530
     :|||::::::      :::|||:::   |||         |||:::   :
 851 eProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSerArgG 868
```

```
531 CCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAG 580
868 lyArgThrIleCysSerSerSerPro.....SerAlaAlaProThr 881

581 CAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACAT 630
882 ProArgAlaArgThrProAlaThrThrProThrProSerSerArgGlnPr 898

631 GAATTTGGCGGCCCGC.....ATCCTGCCGGCTT 659
898 oSerGlySerAlaArgProSerProProSerSerSerAlaIleProArgA 915

660 GAGTGGCACGCACA.....TTCATTTCATCGAGC..... 688
915 rgThrAlaArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAsp 931

689 .....CAGTCGGCGCGAATAAAACCGTG 711
932 SerAlaIleArgArgSerSerThrThrArgSerAlaArgSerArgArgAs 948

712 TGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAAC 761
948 nThrProSerSerAlaSerThr..... 955

762 AGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCA 811
956 .....AlaThrAlaPro...ProThrArgLysPro 964

812 ACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTT 861
965 ThrThrGlySerThrCys.....CysAlaCysAr 974

862 ACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGT 911
974 gProAlaSerThrValAlaAlaArgArgLysLys......ProValArg. 988

912 ATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGG..... 952
989 .....LysValAlaAlaGlnSerSerArgProSerCysTrpLysSerArg 1003

953 ......GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGC 996
1004 SerMetThrAlaThrThrGlyArgThrProThrCysAsnSerAlaArgAr 1020

997 AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCAC 1046
1020 gProValIleSer.....ArgArgSer.....ProSerA 1030

1047 GC.....GCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGT 1090
1030 rgMetPheGlyArgLeuSerAlaSerSerIleAsnMetArgSerThrSer 1046

1091 TCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACT 1140
1047 ValSerAlaProArgThrCys.....ArgAlaThrSe 1057

1141 TATGAGCGCGTAATGCCGTTGGACATCCTGC..... 1171
1057 rSerSerAlaSerCysArgCysLeuSerCysProGlnSerThrThrAlaA 1074

1172 .....CTACCTTGCTTTTGCGCGATTTAATCGTCGGCG 1204
1074 laTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAla 1090

1205 ATACCGACAGCGCGCAGG.....CTTTGGGTTGC 1233
1091 GlyThrThrArgSerArgArgSerSerArgArgThrProSerTrpProSe 1107

1234 TTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGG 1283
1107 rArgAsnTrpTyrSerArgArgArgAsnThrProSerSerAsnSerAlaL 1124

1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGC 1318
1124 ysArgArgArgThrGlyLysValSerArgLysCys 1135
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

ID AAB59827 standard; Protein; 1592 AA.
XX
AC AAB59827;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #4 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1592 AA;

alignment_scores:

Quality:	130.00	Length:	571
Ratio:	0.622	Gaps:	27
Percent Similarity:	36.602	Percent Identity:	22.417

alignment_block:
US-09-303-518D-131 x AAB59827 ..

Align seg 1/1 to: AAB59827 from: 1 to: 1592

```
 32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
620 ProSerGlyAlaSerArgSerSerAlaAlaAlaProArgArgProProPr 636

 82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
636 oSerHisIleSerSerAlaAlaTyrGlyAlaSerCysThr.....A 650
```

```
 132 AATCAAGGAAGGTGAAGCCG..............................  151
       :::|||:::      |||
 650 spAlaArgAsnArgAspProLeuGlyArgGlyCysThrAspProCysCys  666

 152 ............TCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGA..  187
                 :::|||   :::||||||   ::::::   ||| .
 667 ProProAlaThrArgArgLysArgAsnCysCysCysArgAlaAlaArgAr  683

 188 .............ATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAA  224
                 :::|||||| :::      ||||||
 683 gPheArgGlyProLeuArgAla.AlaProAlaCysAlaArgArgPheAla  699

 225 AATCGCCGCTATTCACCG..............................  242
       ||||||   :::|||
 700 TyrArgArgLeuGlyProArgCysThrSerArgGlyArgSerArgCysSe  716

 243 ..................TGGCGAAAAGCGCGTACTTCAGTCAGTCGTG  273
                       |||   ::::::   :::          |||
 716 rProAspArgCysCysArgTrpSerArgCysSerSerProSerProArgA  733

 274 ATTGC.....................................  278
       |||
 733 rgCysProProSerSerProAlaGlyAlaProGlyAlaThrCysSerArg  749

 279 ............CGTTGAAGGCAA............CGACGAAATCGAGT  304
                 |||   |||:::           |||      |||
 750 ArgProPheSerArgSerArgAspSerAlaGlyProArgAlaAlaArgPh  766

 305 TCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTG  354
      |||      |||                    |||:::|||      :
 766 eArgArgCysArg............AspAlaCysGluArgArgAlaArgC  779

 355 CGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCC  404
     ::|||   |||                          |||   ||||||
 779 ysProGlyPro.....................ArgSerAlaProSer  787

 405 GTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGC...........  443
     ::::::::::::         |||   ||||||   |||
 788 IleArgArgGlySerArgAspArgSerArgAlaSerArgSerArgGlySe  804

 444 ................................CATCTTCGTCAATGCG  459
                                     :::   |||:::|||:
 804 rProLeuCysGlyAlaThrAlaThrSerCysProArgArgArgArgCysS  821

 460 ATGGACACCAATCCGCTGGCTGC.....................CGACCC  488
     ::         |||:::||||||                      |||
 821 erIleGlyAlaSerSerGlyCysProHisProProValArgArgSerPro  837

 489 TACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGG  538
       :::      :::|||::::::||||||      :::|||
 838 ValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGlyAr  854

 539 TATTGAGCCGCCT.........GACCGAACGTAAAATCCATGTGTGTAAA  579
      ::::::   |||         |||         :::   |||
 854 gPheArgGlyProThrSerArgAspThrGlyArgArgArgCysTrpArgT  871

 580 GCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACA  629
         |||   ||||||::      :::|||   :::      :::
 871 rpProArgProArgArgCys......ArgCysSerArgArgTrpGlyArg  885

 630 TGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATT  679
       :::|||      :::      |||   :::|||
 886 ProLeuTrpAlaSerGlyCysProArgAlaArgTrp.............  897

 680 TCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAA  729
      :::|||:::|||:::   :::   :::|||   :::      :
 898 .ArgArgGlySerAsnTrpSerSerGlyArgSerSerAlaAlaSerProL  914

 730 GACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGA  779
     ::|||   |||   |||      ||||||:::      |||      |||
 914 ysArgThrCysGlyArgArg...ValArgSerAspThrSerAlaArgArg  929

 780 GCGCGT.....................................  785
```

```
       :::|||
 930 SerArgCysProAlaSerSerProIleArgTrpThrGlyArgCysArgAr  946

 786 ...............GGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCG  819
                    ||||||      |||                 |
 946 gTrpArgArgProLeuGlyCysSerProArg.................A  957

 820 CGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTAC......  863
     ||      :::   |||   |||   :::|||   |||::::::
 957 laThrCysThrAlaArgCysGlyArgAspGlyCysSerAlaPheGly  973

 863 .............................................  863

 974 AsnProLeuHisArgSerLeuArgGlyProTrpAlaAlaProPheArgAl  990

 864 ..................CGCCGGCGAATTGGTTGACGCGGACAACCGC  894
                          ||||||   :::   ||||||:::   |
 990 aHisArgSerArgSerThrThrArgArgCysAlaValArgGlySerSerA 1007

 895 GTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAG.......... 934
     ||      |||   :::      |||  |||   ||||||
1007 rgHisAspArgThrAlaSerThrArg..ArgProHisLysProProLysG 1023

 935 ...................GCGCGCATGATTATTTGGGACGCTACCACA 964
                         :::            |||      |||:::
1023 lyCysAlaThrAspIleHisSerGlyArgTyrCysTrpProArgThrAla 1039

 965 ATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGG 1014
                   |||::::::||||||:::   :::   :::
1040 ..............SerSerArgAlaAlaSerGlyAlaSerAlaLysAr 1051

1015 GTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCA.......... 1054
     |||      |||   :::   |||   |||:::|||
1051 gThrArgLeuArgArgArgSerCysProValArgSerProArgArgArgG 1068

1055 ..CTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
           |||         ::::::   :::|||||||||||||:::|||
1068 lyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArgArgPro 1084

1103 TCAACGGCGGCGACCGCGCCATGG.........TACCGATCGGCACTTAT 1143
     |||::::::         ||||||           |||   :::   ::
1085 SerSerGly......ArgProTrpSerValProIleArgProSerSerIl 1099

1144 GAGCGCGTAATGCCGTTGGA.............................. 1163
     :   ::: :::|||||||||
1099 eCysGly.ArgAlaValGlyLeuThrSerProSerSerProLeuAsnArg 1115

1164 .................CATCCTGCCTACCTTGCTTTTGCGCGATTTAA 1195
                         ||||||                 |||   ||
1116 ProPheAlaArgArgSerAlaProAlaSerThrProCysArgArgHisAs 1132

1196 TCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGAC 1245
     ||||||||||||   ::::|||      |||                 |
1132 nArgArgArgTyrGlySerArgArgProPhe.................A 1143

1246 GAAGAAGA 1253
     ||||||||
1143 rgArgArg 1145
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
```

```
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;


alignment_scores:
           Quality:  129.00              Length:    573
             Ratio:  0.617                Gaps:     27
 Percent Similarity:  36.475   Percent Identity: 22.339

alignment_block:
 US-09-303-518D-131 x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         ||||||   |||:::||||||::::::                ||||||   ||
      25 ProSerGlyAlaSerArgSerSerAlaAlaAlaProArgArgProProPr 41

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
         |:::           |||         ::::::|||:::
      41 oSerHisIleSerSerAlaAlaTyrGlyAlaSerCysThr.........A 55

     132 AATCAAGGAAGGTGAAGCCG............................. 151
           :::|||:::      |||
      55 spAlaArgAsnArgAspProLeuGlyArgGlyCysThrAspProCysCys 71

     152 ............TCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGA.. 187
                     :::|||   :::||||||   ::::::   |||
      72 ProProAlaThrArgArgLysArgAsnCysCysCysArgAlaAlaArgAr 88

     188 ............ATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAA 224
                     :::|||||| :::      ||||||
      88 gPheArgGlyProLeuArgAla.AlaProAlaCysAlaArgArgPheAla 104

     225 AATCGCCGCTATTCACCG............................. 242
           ||||||   :::|||
     105 TyrArgArgLeuGlyProArgCysThrSerArgGlyArgSerArgCysSe 121

     243 ..................TGGCGAAAAGCGCGTACTTCAGTCAGTCGTG 273
                           |||   ::::::   :::         |||

     121 rProAspArgCysCysArgTrpSerArgCysSerSerProSerProArgA 138

     274 ATTGC.......................................... 278
           |||
     138 rgCysProProSerSerProAlaGlyAlaProGlyAlaThrCysSerArg 154

     279 ............CGTTGAAGGCAA............CGACGAAATCGAGT 304
                     |||   |||:::            |||      |||
     155 ArgProPheSerArgSerArgAspSerAlaGlyProArgAlaAlaArgPh 171

     305 TCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTG 354
          |||      |||                        |||:::|||     :
     171 eArgArgCysArg............AspAlaCysGluArgArgAlaArgC 184

     355 CGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCC 404
         ::|||   |||                          |||   ||||||
     184 ysProGlyPro.....................ArgSerAlaProSer 192

     405 GTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGC........... 443
         ::::::::::::         |||   ||||||   |||
     193 IleArgArgGlySerArgAspArgSerArgAlaSerArgSerArgSerAr 209

     444 ....................................CATCTTCGTC 453
                                             :::   |||:
     209 gGlySerProLeuCysGlyAlaThrAlaThrSerCysProArgArgArgA 226

     454 AATGCGATGGACACCAATCCGCTGGCTGC..................... 482
         ::|||::::          |||:::||||||
     226 rgCysSerIleGlyAlaSerSerGlyCysProHisProProValArgArg 242

     483 CGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCC 532
            |||  ::::      :::|||:::::||||||      :::|||
     243 SerProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgAr 259

     533 TGTTGGTATTGAGCCGCCT.........GACCGAACGTAAAATCCATGTG 573
                 ::::::   |||         |||         :::   |||
     259 gGlyArgPheArgGlyProThrSerArgAspThrGlyArgArgArgCysT 276

     574 TGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGA 623
                      |||   ||||||:::      :::|||   :::
     276 rpArgTrpProArgProArgArgCys......ArgCysSerArgArgTrp 290

     624 AACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACA 673
         :::      :::|||     :::       |||  :::|||
     291 GlyArgProLeuTrpAlaSerGlyCysProArgAlaArgTrp........ 304

     674 TTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAAT 723
                :::|||:::|||:::   :::   :::|||   :::
     305 .......ArgArgGlySerAsnTrpSerSerGlyArgSerSerAlaAlaS 319

     724 TATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAA 773
               :::|||   |||   |||      ||||||::      |||
     319 erProLysArgThrCysGlyArgArg...ValArgSerAspThrSerAla 334

     774 TACCGAGCGCGT.................................... 785
           |||:::|||
     335 ArgArgSerArgCysProAlaSerSerProIleArgTrpThrGlyArgCy 351

     786 .....................GGTTGCCTTGGGCGGCCTGCAAGTCAAC 813
                              ||||||      |||
     351 sArgArgTrpArgArgProLeuGlyCysSerProArg............. 363

     814 AAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTAC 863
               |||      :::   |||   |||   :::|||   |||::::::
     364 .....AlaThrCysThrAlaArgCysGlyArgAspGlyCysSerAlaPhe 378

     863 ............................................... 863

     379 PheGlyAsnProLeuHisArgSerLeuArgGlyProTrpAlaAlaProPh 395

     864 .........................CGCCGGCGAATTGGTTGACGCGGAC 888
                                  ||||||   :::   ||||||:
     395 eArgAlaHisArgSerArgSerThrThrArgArgCysAlaValArgGlyS 412
```

```
 889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAG.... 934
 412 erSerArgHisAspArgThrAlaSerThrArg..ArgProHisLysProP 428

 935 ..........................GCGCGCATGATTATTTGGGACGCT 958
 428 roLysGlyCysAlaThrAspIleHisSerGlyArgTyrCysTrpProArg 444

 959 ACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTC 1008
 445 ThrAla..............SerSerArgAlaAlaSerGlyAlaSerAl 456

1009 GGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCA.... 1054
 456 aLysArgThrArgLeuArgArgArgSerCysProValArgSerProArgA 473

1055 ........CTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGA 1096
 473 rgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArg 489

1097 CAGCCGTCAACGGCGGCGACCGCGCCATGG.........TACCGATCGGC 1137
 490 ArgProSerSerGly......ArgProTrpSerValProIleArgProSe 504

1138 ACTTATGAGCGCGTAATGCCGTTGGA.......................... 1163
 504 rSerIleCysGly.ArgAlaValGlyLeuThrSerProSerSerProLeu 520

1164 ........................CATCCTGCCTACCTTGCTTTTGCGCG 1189
 521 AsnArgProPheAlaArgArgSerAlaProAlaSerThrProCysArgAr 537

1190 ATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAA 1239
 537 gHisAsnArgArgArgTyrGlySerArgArgProPhe.............. 549

1240 TTGGACGAAGAAGA 1253
      |||||||||
 550 .....ArgArgArg 552
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998

seq_documentation_block:

```
ID   AAY04998 standard; Protein; 388 AA.
XX
AC   AAY04998;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 50B.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34249.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 50B; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   388 AA;
```

alignment_scores:

```
            Quality:  127.50            Length:    471
              Ratio:  0.750               Gaps:     27
 Percent Similarity:  36.093   Percent Identity: 22.718
```

alignment_block:

US-09-303-518D-131 x AAY04998 ..

Align seg 1/1 to: AAY04998 from: 1 to: 388

```
  72 GGCCATTACCGAAGTCGCGTTGCTTGG......CGAAGAATATGTCGGCA 115
  25 GlyArgArgArgArgArgTrpGlnTrpArgGluArgArgArgCysGlnHi 41

 116 T.....................................................G 117
  41 sArgGlnArgTrpArgArgTrpProArgArgCysArgTrpArgTrpArgM 58

 118 CGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGT 167
  58 etAlaLeuArgArgArgArgGlyArgArgThrArgArgThrArgArgAsn 74

 168 GCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTT 217
  75 ArgProArgArgArgArgArgArgArgArgGly.........ProGlyAr 88

 218 CAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTACT....... 260
  88 gArgArgProArgThrValGlyTyrTrpArgArgArgArgThrArgArgA 105

 261 .......................TCAGTCAGTCGTGATTGCCGT...... 281
 105 laArgArgTrpTyrArgGlyProThrAlaAlaArgSerGlyArgHisGly 121

 282 ........................TGAAGGCAACGACGAAATCGAGTTCG 307
 122 ArgArgGlyTrpArgArgTrpAlaAspArgGlnArgArgGlyArgArgAr 138

 308 AACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGC 357
 138 g..................................................... 138

 358 CGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTT 407
 138 ..................................................... 138

 408 CAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATG 457
 139 ..............ArgArgCysArgArgValArgArg............ 146

 458 CGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAA 507
 147 ....GlyArgArgSerArgArgCysArgArgGluArgHisAlaAspArgA 162

 508 GCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGA 557
```

```
  162 laArgArgArgArgArg.............................. 167

  558 ACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAA 607
                                 ||||||||||||
  168 ..........................ArgArgArgArgGln........ 172

  608 ATGCTGCCAATATCGAAACACATGAATTTGG.........CGGCCCGCAT 648
          |||:::|||            |||          |||   :::
  173 .......GlnPheArg.........***TrpArgGlyArgArgArgGlyA 184

  649 CCTGCCGG.............CTTGAGTGGCACGCACATTCATTTCATCGA 686
        ||||||                :::|||::::::             |||
  184 rgCysArgArgAlaProLeuArgGlnTrpArgVal............Arg 196

  687 GCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGA 736
            |||||||::   :::      :::|||:::
  197 ArgProArgArgSerArgHisGlyArgGlnHisArg.............. 208

  737 TTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTG 786
          ||||||                                :::|||   |||
  209 ....TyrArg..........................TrpArgArgArgA 215

  787 GTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGT 836
              ||||||   |||:::::::
  215 rgArgTrpGlyArgGlnAlaAspArg....................... 223

  837 TTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGG 886
                                 :::|||||||||          |||||
  224 .................ProArgTrpArgArgArgCysArg...ArgGl 233

  887 ACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGC 936
      |      |||::::::   :::                    :::|||||
  233 yProArgArgSerLeuGlyTrpPro.................***ArgA 244

  937 GCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGA 986
      |||||                                  |||::::::|||
  244 rgAla..............................ArgTrpGluArg 249

  987 AGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAAT 1036
                :::   |||:::|||
  250 GlyArgTrpArgProAlaIleArg........................... 257

 1037 ACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTC 1086

  257 ................................................. 257

 1087 AAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGG 1136
                  :::|||:::|||||||||||||||||||
  258 ...........GlnArgArgArgArgArgProArgHis............ 266

 1137 CACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTT......GC 1180
                   ||||||::::::|||           :::      ::
  267 .........ArgArgAsnThrAlaGlyGlyGlyGluArgIleGlyAspGl 280

 1181 TTTTGCGCGATTTAATCGT......CGGCGATACCGACAG.......... 1214
      :|||:::|||   :::|||      ||||||:::|||
  280 yPheValArgCysThrArgProThrArgArgHisArgProThrArgLeuA 297

 1215 ........CGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCT 1256
             :::   |||||||||      |||   |||:::   :::|||
  297 laProIleAsnGluGlyPheGly...AlaGlyProGlyHisGlyHisPro 312

 1257 ....................CGCTTTGTGCAGCTTCGTCTG...CCCGG 1282
                          |||:::   :::      |||   |||||
  313 LeuSerTrpArgSerGlyLeuArgTrpCysArgLysIleLeuProProGl 329

 1283 GCAAATACGAATA 1295
      |   :::|||:::
  329 yProLeuArgLeu 333
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146

seq_documentation_block:

```
ID   ABG05146 standard; Protein; 440 AA.
XX
AC   ABG05146;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5137.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69333.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 35505; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   440 AA;
```

```
alignment_scores:
            Quality:  125.00              Length:    480
              Ratio:  0.654                 Gaps:     26
 Percent Similarity:  39.792   Percent Identity: 22.917

alignment_block:
 US-09-303-518D-131 x ABG05146   ..

 Align seg 1/1  to: ABG05146  from: 1  to: 440

      56 TCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAA 105
         ||||||::::::   :::|||   |||::::::|||      :::   ::
```

```
 28 SerPheLeuSerProGlnProProProArgAlaArgSerSerSerAlaTh 44

106 TATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAA 155
    :::|||:::   |||||||||         |||   :::   ||| |||
 44 rLeuSerSerGlyAlaProArg.........ArgAlaAlaThrPro.Gln 57

156 AAAAGG.........CCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCG 196
      :::         ||||||:::|||            :::
 58 ProGlnThrThrProProSerSerVal............GlnProCysAl 70

197 TAGTATTTACTGCGCC...GGCTTCAGGCAAAATCGCCGCTATTCACCGT 243
     :::      ||||||   |||   |||                      |
 70 aAlaThrProCysAlaArgGlyGlyArgProAlaProAlaGlyValThrT 87

244 GGCGAAAAGCGCGTACTTCAGTC.......................... 266
    ||      :::|||   |||:::
 87 rpProThrThrArgProSerMetAla***ArgThrSer***AlaSerThr 103

267 ..........................AGTCGTGATTGCCGTTGAAGGCA 289
                               |||:::   |||        ::
104 ThrAlaProArgMetAlaProProArgSerHisAlaCysAlaArgSerHi 120

290 ACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTG 339
    :|||   :::   :::|||:::   ||||||         :::   |||
120 sArgProGluThrAlaArgSerAlaArgThrAlaProArgSerAlaIleT 137

340 AGCAGCGA.........AAAAGTGCGCCGCAACCTGATTCAATCAGGCTT 380
      :::|||         ::::::   |||   |||   :::      :::
137 hrArgArgAlaPheThrSerThrArgProProProThrThrArgThrVal 153

381 ATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATG 430
     :::   ::::::|||   |||      :::      :::   |||
154 AlaSerSerGlyThrHisThrSerGlyLeuSerProThrAlaSerArgLe 170

431 CCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480
                                                |||   |
170 u.....................................AlaArgC 173

481 GCCGACCCTACGGTCATCATCAA.....................AGAAGC 509
    |||||      |||:::   :::                    ||||||
173 ysArgAlaProGlyArgSerSerThrIleIleThr***ThrCysArgSer 189

510 CGCCGAAGACTT.........CAAACGCGGCCTGTTGGTATTGAGCCGCC 550
          ||||||                 ||||||:::         |||||
190 ProThrArgLeuCysCysProAlaGlnArgProLeu.........ProPr 203

551 TGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600
    |     :::   :::            |||||||||      |||:::|
203 oAlaSerSerProSer............SerSerArgThrSerArgSerV 216

601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCC 650
    ||                                  |||
216 al..............................TrpThrArgArgCys 221

651 TGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGA 700
      ||||||:::|||
222 ThrArgLeuArgTrp................................ 226

701 ATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGT 750

226 ..................................................... 226

751 TTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGG 800
    :::   :::||||||   ::::::   |||               |||
227 ..ThrSerSerArgProProSerTrpMetAlaLeu............Arg 238

801 CCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGG 850
      ::::::   :::|||:::|||   |||   |||     :::||||||
239 ThrValGlyThrSerThrGlyProThrAla***ArgSer..LeuArgArg 254

851 TGT............................CTCAACTTACCGCCGGC 870
    |||                           :::   ||||||
255 CysGlnAlaSerThrTrpArgSerArgProSerThrSerAlaProPro.. 270

871 GAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGG 920
      :::|||   :::|||   :::|||      |||:::
271 .SerTrpCysAlaArgTrpAlaAlaThr***ProLeu.........ProS 284

921 TGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGA 970
              ::::::         ::::::|||   :::   |||   |||
284 erAlaCysGlnArgLysTrpSerMetLeuTrpArgThrGlyThrAlaArg 300

971 TTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020
              |||        :::|||||||   |||   |||   :::
301 ValSerThrSerAlaCysGlyAlaAlaProSerThrSerArgSer..... 315

1021 CCGCAGCCGGACAAATACTCCATCA...CGCGCACCACTCTCGGCCATTT 1067
       ||||||   :::|||||||      ||||||||||:::   |||   :
316 .ThrSerArgProSerThrProMetLeuArgAlaProValProAla...G 331

1068 CCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACC 1117
     ::   :::      :::         :::   |||      :::   |||
331 lyTrpGlnProProAlaLeuHisProGlnProProArgProSerHisThr 347

1118 GCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATC 1167
               :::   ||||||     |||      |||||||||   ::
348 ArgGlnProTrpProSerAlaArgArgSer......CysArgTrpArgTh 362

1168 CTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGC 1217
     :|||              ::::::      ||||||         |||:::
362 rCysThrThrArgProAlaSerSerThrSerSerProArgAlaThr***T 379

1218 GCAGGCTTTGG.........GTTGCTTGGAATTGG............... 1243
               |||         :::   |||:::
379 hrSerHisTrpProProThrThrArgTrpArgMetSerArgCysSerThr 395

1244 ...ACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATAC 1290
        ||||||   :::      |||      :::   :::   ::::::::
396 ProThrLysThrAsnCysThrCysMetArgGlyLeuGlyThrCysGlnAl 412

1291 GAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCA 1327
     ::::      |||  |||         |||   |||
412 aGlyArgLeuArgGlyCysPro......TrpProPro 422
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955

seq_documentation_block:

```
ID   AAY04955 standard; Protein; 573 AA.
XX
AC   AAY04955;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 41T#3.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34206.
```

XX
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
XX
PS Claim 32; Fig 41T; 309pp; French.
XX
CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC proteins from various Mycobacterium species microorganisms. The
CC encoding nucleotide sequences can be used as primers and probes for
CC methods for detecting and identifying mycobacteria, especially belonging
CC to the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
XX
SQ Sequence 573 AA;

alignment_scores:
Quality: 119.00 Length: 477
Ratio: 0.626 Gaps: 20
Percent Similarity: 39.832 Percent Identity: 23.270

alignment_block:
US-09-303-518D-131 x AAY04955 ..

Align seg 1/1 to: AAY04955 from: 1 to: 573

```
 72 GGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCC 121
    |||:::||||||      :::|||   ||||||      |||:::
166 GlyAsnTyrArgLeuGlyAlaAlaGlyArgArgArgSerArgArgProVa 182

122 CCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGG.......... 161
          ::::::   ||||||   ::::::::::::|||
182 lArgAlaArgGlyValGlyArgCysGlyHisArgArgArg***ArgGlyG 199

162 ...................CCAAGTGCTGTTTGAAGACAAAAAGAATCC 191
                        |||   ::::::   |||   ::::::
199 lyHisArgAlaGlyLysAspProArgThrAla***ArgAlaArgArgCys 215

192 GGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACC 241
    ||||||:::                    :::||||||   :::||
216 GlyArgGly.....................GlyArgArgArgThrGlyPr 225

242 GTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAAC 291
    |      :::|||         :::   :::      |||
225 oAlaGlySerAlaGlyArgValAlaLeuHisHisLeuArgAlaGlyThrC 242

292 GACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAG 341
          :::   :::||||||||||          |||:::::::::
242 ysProGlyGlyLeuArgThrLeu............AlaAlaArgValAla 254

342 CAGCGAAAAAGTGCGCCGCAACCTGAT.....................TC 370
    :::|||   :::   |||   |||                        ||
255 AspArgHisGlyTyrProThrProArgProAlaIleArgGlyAspValSe 271

371 AATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCT 420
    |:::      |||   |||:::|||         |||      :::
271 rValGlyGlyMet***CysCysSerGlyGlyProVal...AlaGlySerT 287

421 GCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAA 470
      :::        :::|||         |||:::|||:::::::::|||
287 hrGlnGlyIleGly***ValGlyGlyHisArgArgCysSerAlaArgGln 303

471 TCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACT 520
                  ||||||      |||:::::::|||      ||||||
304 LeuLeuArgThrArgPro......HisArgArgArgArgCysArgArgGl 318

521 TCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
     :::           |||   :::   |||||||||      :::
318 ySerArgIleGlyGlyGlyAlaSer***ProAspArgAspLeuGly.... 333

571 GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATAT 620
                                                :::::::::
```

```
334 .....................................AlaArgPhe 336

621 CGAAACACATGAATT.........TGGCGGCCCGCATCCTGCCGGCTTGA 661
    |||:::      |||         |||      |||:::   |||
337 ArgAspGlnArgIleAlaGlyArgTrpLeuAspAlaGlyProArg..... 351

662 GTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTG 711
                            |||||||:::   |||:::
352 .....................ArgAlaGlyGlyArgArg.......... 357

712 TGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAAC 761
                      ||||||   |||   |||:::
358 .................ArgArgArgCysArgArgAla............ 364

762 AGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCA 811
                    |||       |||   |||          ::::::::
365 ............ValArgArgGlyGlyArgLeuArgAlaAlaThrGlySe 377

812 ACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGC............GAAG 849
    ::::            :::   |||:::   |||              |
377 rArgArgArgAspThrGlyArgArgTyrGlnCysProProAlaGlyAlaG 394

850 GTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGAT 899
    ||   :::   :::|||||||||
394 lyArgGlyArgHisArgArgArgAla........................ 402

900 TTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATT 949
       |||   |||       :::       |||   ||||||
403 ...ArgAspGlyAlaAlaGlnTrpLeuCysGlyArgArgArg...ThrGl 417

950 TGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAA 999
     |||   :::   :::::::|||   |||      ||||||         |
417 yGlyArgValTyrArgGlyAsp...ArgLeuGlyArgArgArgGlyThrA 433

1000 GAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA............ 1037
     |||||   |||:::      |||:::::::|||:::
433 rgAlaAspArgIleAspGlyAlaGlyValGlyArgAlaGlyArgAla*** 449

1038 ..........................CTCCATCACGCGCACCACTCTCG 1060
                                 |||:::|||:::         ||
450 ArgGlyProProGlyArgArgArgArgLeuGlnHisGlyProCysArgAr 466

1061 GCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGC.......... 1100
     |   ||||||   :::         :::   |||   :::
466 gCysPheProAlaArgIleGlyAlaHisCysHisProLysGlyAlaAspL 483

1101 ..........................CGTCAACGGCGGCGACCGCGCCAT 1124
                                 |||:::      |||   |||:::
483 euGlyArgTyrLeuGlnAlaGlyArgArgSerGlyTyrArgGlyArgArg 499

1125 GGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTA 1174
     |||:::||||||
500 GlyAlaAspArg..................................... 503

1175 CCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAG.........C 1215
                              |||||||||   |||:::         |
504 .....................ArgArgArgCysArgArgGlyGlyHisA 513

1216 GCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTG 1265
     ||:::|||      ::::::||||||||||||||||             :::
513 rgSerGlyArgProValValGlyIleGlyArgArgSerGlyAspGlyAla 529

1266 CAGCTTCGT.......................CTGCCCGGGCAAATACG 1291
       :::   |||                              ||||||
530 AsnTrpArgArgArgAsnArgArgArgGlyCysArgProGlyThrAlaCy 546

1292 AATACGGCCCGCTGTTGCGCAAAGTGCTGGA 1322
         ||||||         :::   ||||||
546 sAlaArgProProSerArgHisArgAlaGly 556
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935

seq_documentation_block:

```
ID   AAB15935 standard; Protein; 740 AA.
XX
AC   AAB15935;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated protein sequence SEQ ID NO:292.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   N-PSDB; AAA65940.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 11; Page 217-219; 316pp; English.
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
SQ   Sequence   740 AA;


alignment_scores:
           Quality:  117.00              Length:    313
             Ratio:  0.731                Gaps:     13
 Percent Similarity:  51.118   Percent Identity: 22.684

alignment_block:
 US-09-303-518D-131 x AAB15935   ..

 Align seg 1/1  to: AAB15935  from: 1  to: 740

     100 GAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGC 149
         :::::::::::::|||   :::  ::::::   ::::::   |||:::
      48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64

     150 CGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAG 199
          |||   :::||||||   |||   :::   :::   :::         |
      64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuProV 81

     200 TATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATT..... 237
         ||      ||||||:::||||||   ::::::||||||
      81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96

     238 ......CACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
               |||   :::            |||   ||||||:::|||   ::
      97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAA..... 324
         ::::|||:::|||                  ::::::|||:::
     112 aAspGlyGluAsp.............CysTrpIleProArgAspGlyT 124

     325 .....GCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
                          |||   |||::::::   :::::::::
     124 rpAlaAspTyrArgThrArgSerArgGluGluLeuIleGluArgIleHis 140

     370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGC..... 411
         |||   |||:::   ::::::|||           |||
     141 GlnPheGlyVal...AlaGlyLeuGlyGlyAlaGlyPheProThrGlyVa 156

     412 .AAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGA 460
          |||:::   :::         :::      ::::::::::::||||||
     156 lLysLeuGlnGlyGlyGlyAspLysIleGluThrLeuIleIleAsnAlaA 173

     461 TGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCC 510
           :::   :::|||       :::              ::::::::::::::
     173 laGluCysGluProTyrIleThrAlaAspAspArgLeuMetGlnAspCys 189

     511 GCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAA.. 558
         |||   :::     :::|||:::   :::|||:::::::::   :::
     190 AlaAlaGlnValvalGluGlyIleArgIleLeuAlaHisIleLeuGlnPr 206

     559 ..................................CGTAAAATCCATG 571
                                          :::   |||   :
     206 oArgGluIleLeuIleGlyIleGluAspAsnLysProGlnAlaIleSerM 223

     572 TGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATC 621
         ::   :::|||:::   ||||||              :::   :::|||
     223 etLeuArgAlaValLeuAlaAsp..............SerAsnAspIle 234

     622 GAAACACATGAATTTGGCGGCCCGCATCCTGCCGGC.............. 657
         :::   :::   :::         :::|||:::|||
     235 SerLeuArgValIleProThrLysTyrProSerGlyGlyAlaLysGlnLe 251

     658 .........TTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCG 697
                 |||:::|||   ::::::            |||   |||:
     251 uThrTyrIleLeuThrGlyLysGlnVal............ProHisGlyG 264

     698 CGAATAAAACCGTGTGGACCATCAATTATCAAGACGTG.........ATT 738
         ::::::::::          :::      |||:::|||
     264 lyArgSerSerAspIleGlyValLeuMetGlnAsnValGlyThrAlaTyr 280

     739 GCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGT 788
         |||:::   |||       :::   |||:::       |||||||||||||
     281 AlaValLysArgAlaValIleAspGlyGluProIleThrGluArgValVa 297

     789 TGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTT 838
         |:::|||   |||       :::   :::|||       :::   :::   |
     297 lThrLeuThrGlyGluAlaIleAlaArgProGlyAsnValTrpAlaArgL 314

     839 TGGGTGCGAAGGTGTCTCAACTT......ACCGCCGGCGAATTGGTTGAC 882
         |||||:::   |||   :::|||          ||||||          :::
     314 euGlyThrProValArgHisLeuLeuAsnAspAlaGlyPheCysProSer 330

     883 GCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGT 921
         ||||||:::   ||||||   |||:::   |||   |||
     331 AlaAspGlnMetValIleMetGlyGlyProLeuMetGly 343

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731

seq_documentation_block:
ID   ABG03731 standard; Protein; 696 AA.
XX
AC   ABG03731;
XX
DT   13-FEB-2002  (first entry)
```

```
XX
DE   Novel human diagnostic protein #3722.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS67918.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 34090; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   696 AA;


alignment_scores:
            Quality:  115.50              Length:    462
              Ratio:  0.553                Gaps:     23
 Percent Similarity:  45.238   Percent Identity: 24.459

alignment_block:
 US-09-303-518D-131 x ABG03731   ..

 Align seg 1/1  to: ABG03731  from: 1  to: 696

      27 TCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCA 76
         ||||||  ||||||:::      |||           :::   |||
       9 SerAla***ArgGlyAsnLeuArgAlaArgLysProSerLysArgGlyLy 25

      77 TTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC. 125
            |||:::|||                    |||   :::|||
      25 sAspArg***Arg.....................ArgValSerProThrA 35

     126 .....................GATGAAAATCAAGGAAGGTGAAGCCGTCAA 155
                              |||::::::::::|||   :::::::::
      35 rgSerGlyLysArgArgGlyAlaGluGluLysAsnArgGlnGluLysLys 51

     156 AAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTA 205
         |||      :::        |||:::|||:::|||   |||:::
      52 LysGlyArgGluLysGluArgArgGluLysArgSerGluArgGln..... 66

     206 CTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGC 255
                 |||:::::::||||||   :::      ||||||
      67 .............ArgAspArgArgArgArgLysGluGlnArgLys.... 77

     256 GTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTT 305
                    ::::::   |||   |||   |||   |||:::
      78 ..............GluGluGlnArgArgArgAlaArgThrAsnGluArg 89

     306 CGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGC 355
         :::      :::|||   :::   |||      ::::::   |||:::::
      90 LysProArgGlnThrGlnAlaAsnGlyAlaThrSerSer***LysAlaSe 106

     356 GCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCG 405
         :   |||      ::::::                ::::::   |||::::
     106 rAlaGlnGlnAlaGlyMet............TrpGlyGlySerPro***T 119

     406 TTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAA 455
         :::::   :::      ||||||      |||            |||:::
     119 hrAspAlaThrAlaIleArgArgGlyGlyAlaProCysSerSerArgArg 135

     456 TGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAG 505
                 :::|||::::::   :::   |||   |||:::
     136 ThrCysLeuAsnGlnGlyThrIleAlaThrProSerGlyArg........ 149

     506 AAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACC 555
         |||||||::   ::::::      ||| |||:::
     150 ....ArgArgHisGlyAspAlaGly***ProGly.LeuAlaSerGluHis 164

     556 GAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGA 605
         :::   :::   |||   |||      :::||||||      ||||||::
     165 AspAlaSerGlyHisGlyCysLeuArgThrGlyAlaGly***ProSerAs 181

     606 AAAT..........................GCTGCCAATATCGAAACAC 628
         ::::                          |||   ::::::   ||||
     181 pSerThrGluSerValCysArgArgProLeuAlaMetHisValProThrH 198

     629 ATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCAT 678
         |||||     |||||| ::::::         |||::::::||||||
     198 isGluSerHisGlyPro.ValPheThrArgLeuValSerHisThrPheHi 214

     679 TTCATCGAGCCAGTCGGCGCGAATA...AAACCG......TGTGGACCAT 719
             :::|||:::        :::   :::|||      |||   |||:
     214 sCysGlySerLysLeuProAlaValGlyArgProValAlaCysArgProT 231

     720 CAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTC 769
         ::    |||         |||::::::            :::||||||::
     231 hr.TyrSerProSerLeuCysHisAsnPro.........GlnArgProAl 244

     770 TGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCG 819
         ::::                  |||::::::   ::::::   |||   |
     244 aGln..............LeuLeuAlaHisSerSerAlaLeuGlnCysA 256

     820 CGCCTCTTGCG..................................... 830
         |||||||||:::
     256 laProLeuSerTrpAspProGlnArgCysAlaProProSerProArgPro 272

     831 TACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTG 880
         :::|||   |||      :::         :::||||||
     273 HisArgArgGlyProProSerProHisProHisArgArgAla........ 286

     881 ACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCA 930
             :::|||:::      |||            |||:::      |
     287 .ProProSerProHisProHisArg............ArgAlaHisThrT 299
```

```
 931 CAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTAT 980
     ||   |||               |||   |||   ::::::
 299 hrAlaArg.............ThrAspProThrThrSerAlaProPro 310

 981 CGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGG 1030
        |||:::   |||:::||||||:::|||                |||
 311 ProArgGlnThrGlnArgArgAlaThrArg.........GluProAlaTh 324

1031 ACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAA 1080
      :::      ::::::||||||      |||         ::::::   |
 324 rLysHisThrArgAsnAlaHisProArgArgSerAlaCysAsnArgGlyT 341

1081 CTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACC 1130
     ||         |||         :::||||||||||      :::   |||
 341 hr.........HisThrHisProArgArgArgArgThrThrGluArgThr 354

1131 GATCGGCACTTATGAGCGCGTAAT.......................... 1154
        :::|||         ||||||
 355 ThrHisHisAlaArgProArgAsnArgGlyGlnAlaThrProAsnThrAr 371

1155 ..........GCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTA 1194
               |||      |||          :::   :::|||
 371 gGlnProThrAlaGlyArgHisGluGluThrAspGlyAlaThrArg.... 386

1195 ATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGA 1244
       ||||||::::::   |||      |||   |||            |||
 387 ..ArgArgGlnHisGlyGlnThrArgGlyGluGly...........Gly 398

1245 CGAAGAAGACCTCGCTTTGTGCAGCTTCGT 1274
     |||||||||   |||   ::::::   |||
 399 ArgArgArgGlyArgAlaAlaLysThrArg 408
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820

seq_documentation_block:

ID AAG90820 standard; Protein; 1209 AA.
XX
AC AAG90820;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4574.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66039.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 4574; 246pp + Sequence Listing; English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1209 AA;

alignment_scores:

Quality:	113.50	Length:	554
Ratio:	0.467	Gaps:	28
Percent Similarity:	43.863	Percent Identity:	21.480

alignment_block:
US-09-303-518D-131 x AAG90820 ..

Align seg 1/1 to: AAG90820 from: 1 to: 1209

```
  97 GGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAA...ATCAAGGAAGG 143
     |||:::      |||:::::::|||   :::::::|||   ::::::::::||
 664 GlyLysThrGluValAlaValArgAlaAlaPheLysAlaValGlnAspGl 680

 144 TGAAGCCGTC.............................AAAAAAGGCC 163
     |:::   |||                                ::::::   |
 680 yLysGlnValAlaValLeuValProThrThrLeuLeuAlaGlnGlnHisG 697

 164 AAGTGCTGTTTGAAGACAAAAAGAAT......CCGGGCGTAGTA...... 201
     ||      ||||||::::::   :::      |||   ::::::
 697 lnSerThrPheGluGluArgMetThrGlyPheProValThrIleLysGly 713

 202 .........TTTACTGCGCCGGCTTCAGGCAAA..............AT 227
             ||||||:::||||||:::::::::              ::
 714 LeuSerArgPheThrSerProAlaGluSerArgGluIleLeuSerGlyLe 730

 228 CGCCGCT............ATTCACCGTGGCGAAAAGCGCGTACTTCAGT 265
     :||||||            |||      |||      |||:::||||||:
 730 uAlaAlaGlySerValAspIleValIleGlyThrHisArgLeuLeuGlnT 747

 266 CA.........................GTCGTGATTGCCGTTGAAGGC 288
     ::                        |||::::::      |||
 747 hrGlyValGlnTrpLysAsnLeuGlyLeuValIleValAspGluGluGln 763

 289 AACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCG........... 327
     :::      :::|||   ::::::::::::
 764 ArgPheGlyValGluHisLysGluHisIleLysAlaLeuArgThrHisVa 780

 328 ....CTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAAT 373
         :::      :::|||:::      ::::  |||:::|||   :::|
 780 lAspValLeuThrMetSerAlaThrProIleProArgThrLeuGluMetS 797

 374 CA......GGCTTA......TGGACTGCGCTTCGCACCCGTCCGTTCAGC 411
     ||      |||:::        |||::::::   |||   |||   :::
 797 erMetAlaGlyIleArgGluMetThrThrMetLeuThrProProGluAsp 813

 412 AAAATCCCTGCCGTA................................. 426
     :::   |||   :::
 814 ArgHisProIleLeuThrTyrValGlyProTyrGluAspLysGlnValAl 830

 427 ............................GATGCCGAGCCGTTCGCCATCT 448
                                |||::::::   |||   |||
 830 aAlaSerIleArgArgGluLeuLeuArgAspGlyGlnValPhePheIleH 847

 449 TCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATC 498
         ::::::   |||   :::      |||   :::         :::
```

```
 847 isAsnLysValAlaAspIleGluLysLysAlaArgGluIleArgAspLeu 863

 499 ATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCG 548
     :::   ||||||                 :::|||   :::      :::
 864 ValProGluAlaArgValValValAlaHisGlyGlnMetSerGluGluLe 880

 549 CCTG.........................ACCGAACGTAAAATCCATG 571
      |||                         |||      :::   |
 880 uLeuGluGlnThrValGlnGlyPheTrpAspArgGluTyrAspValLeuV 897

 572 TGTGTAAAGCA.........GCAGGCGCAGACGTGCCGTCTGAAAATGCT 612
     |||||   :::         :::|||   |||:::   :::   |||:::
 897 alCysThrThrIleValGluThrGlyLeuAspIleSerAsnAlaAsnThr 913

 613 GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAG 662
           :::|||                :::   |||      ||||||||
 914 LeuIleValGlu..............AsnAlaHisHisMetGlyLeuSe 925

 663 TGGCACGCAC..................................... 672
     |     |||
 925 rGlnLeuHisGlnLeuArgGlyArgValGlyArgSerArgGluArgGlyT 942

 673 ..ATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720
          :::|||:::   |||   ||||||      |||:::      :::
 942 yrAlaTyrPheLeuTyrProLysGlyAla......ThrLeuThrGluMet 956

 721 AATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCT 770
     :::|||:::   :::   :::|||:::::         :::::   ||
 957 SerTyrAspArgLeuAlaThrIleAlaGln.........AsnAsnAspLe 970

 771 GAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGC 820
     |:::::      :::|||||::      |||:::::          |
 970 uGlyAlaGlyMetAlaValAlaMetLysAspLeuGluMet.........A 984

 821 GCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCC... 867
     ||        :::|||||||||||:::   |||         |||
 984 rgGlyAlaGlyAsnValLeuGlyAlaGluGlnSerGlyHisIleAlaGly 1000

 868 ............................GGCGAATTGGTTGACGCGGA 887
                                 ||||||   |||:::|||
1001 ValGlyPheAspLeuTyrValArgLeuValGlyGluAlaValGluAlaTy 1017

 888 CAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATT.......... 927
      :::   :::   :::|||:::|||:::::|||:::::
1017 rArgAlaLeuAlaAspGlyLysValValAspGlyThrValLysGlyProL 1034

 928 ............................GCACAAGGCGCGCATGATTAT 948
                                 |||:::      :::   |||
1034 ysGluIleArgValAspLeuProValAspAlaHisIleProGluLysTyr 1050

 949 TTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAA 998
     ::: :::|||      :::|||:::   :::   :::|||   |||   :
1051 Ile.AsnAlaGluArgLeuArgLeuGluIleTyrArgLysLeuAlaGlnS 1067

 999 AGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGC 1048
     :::::   ::::::   ||||||    |||   :::   ::::::
1067 erGluSerGluValAspLeuArg.LeuAlaValGluGluMetGluAspAr 1083

1049 GCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACA 1098
      :::         |||         ::::::   |||   |||
1083 gTyrGlyProIleProGluGluValGluArgLeuLeuAlaValSerArgL 1100

1099 GCCGTCAACGGCGGCGACCGCGCCATGGTACC............GATCGG 1136
       |||:::      |||      |||   |||           ::: ||
1100 euArgHisLeuMetArgGluAlaHisLeuThrAspIleAlaValGln.Gl 1116

1137 CACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACC......TTGC 1180
     ||||      :::|||   |||:::|||:::      :::      :::
1116 yThrArgIleLysValHisProValAspLeuAlaAspSerGlnGlnValA 1133

1181 TTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGT 1230
       |||:::   |||::   |||   |||      |||   |||
1133 rgLeuLysArgLeuPheProGlyAlaThrTyrArgAla...AlaAlaLys 1148

1231 TGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCC 1280
     ::::::::::|||                          ||||||
1149 AlaIleGlnLeu.......................SerPheProLysTh 1157

1281 GGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTG............. 1317
      |||:::          ||||||||||||   |||
1157 rGlyAsnLysValThrAspProLeuLeuArgAspValAspLeuLeuGlnT 1174

1318 ...............................CTGGAAACCATTGAGAAG 1335
                                     ||||||   |||:::
1174 rpValAlaAsnPheIleSerAsnMetPheAsnLeuGluGluIleAspVal 1190

1336 GAAGGC 1341
     :::|||
1191 ArgGly 1192
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB96063

seq_documentation_block:

```
ID   AAB96063 standard; Protein; 822 AA.
XX
AC   AAB96063;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Putative P. abyssi pyruvate kinase #1.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS   Pyrococcus abyssi.
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 7; Pages 677-680; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC   a hyperthermophilic archaeon, which is isolated from deep-sea
CC   hydrothermal vents. The present sequence is one such P. abyssi protein.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence   822 AA;
```

```
alignment_scores:
            Quality:  113.00              Length:    475
              Ratio:  0.467                 Gaps:     20
 Percent Similarity:  50.947   Percent Identity: 20.421

alignment_block:
```

US-09-303-518D-131 x AAB96063 ..

Align seq 1/1 to: AAB96063 from: 1 to: 822

```
 31 CCCATCGCGGGCAGACCGGAGCAAGTCATTTATGAC..............  66
238 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 254

 67 .GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCA 115
254 uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 271

116 TGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAA 165
271 luLysGlyThrTrpLysIleLysGlu...LysValIleAlaLysLysGlu 286

166 GTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGC 215
287 ValMet.........................ValIleArgAsnProGl 294

216 TTCAGGCAAA..............ATCGCCGCTATTCACCGTGGC.... 246
294 uThrGlyLysGlyThrValThrValLysValAlaGluTyrLeuGlyProG 311

247 ........GAAAAGCGCGTACTT.........CAGTCAGTCGTGATTGCC 279
311 luTrpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAla 327

280 GTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCT 329
328 LysMetGlyGlnLysIleGluGluHisTyrGlyTrpProGlnAspIleGl 344

330 GGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCT 379
344 uTrpAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgP 361

380 TATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGAT 429
361 roValThrThrLeuLysGluThrThrThrGluGluValGluGluVal... 376

430 GCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG...CT 476
377 ...GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIl 392

477 GGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAAC 526
392 eGlyAlaGlyArgValValValIlePheAspAlaSerGlu.......... 405

527 GCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGT 576
406 ..............IleAspLysValLysGluGlyAspValLeuValThr 417

577 AAAGCAGCAGGCGCAGAC...GTGCCGTCT...GAAAATGCTGCCAATAT 620
418 ThrMetThrAsnProAspMetValProAlaMetLysArgAlaAlaAlaIl 434

621 CGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGC 670
434 eIleThrAspGluGlyGlyArgThrSerHisAlaAlaIleValSerArgG 451

671 ACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720
451 luLeuGlyIleProAlaValValGlyThrLysGluAlaThrLysLysLeu 467

721 AATTATCAAGACGTGATTGCTATCGGA..................... 747
468 LysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuValTyrLy 484

748 ..........CGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGG 787
484 sGlyIleValLysSerLeuValGluLysLysLysLysGluGluAlaAlaA 501

788 TTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTT 837

501 laAlaProGlyAlaAlaValAlaAlaAlaProLeuValThrGlyThrLeu 517

838 TTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGA 887
518 ValLysValAsnValSerMetProGluValAlaGlu.............. 529

888 CAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCG 937
530 ....ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaG 545

938 CGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAA 987
545 luHisMetIleLeuSerIleGlyGlnHisProValLysPheIleLysGlu 561

988 GGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATA 1037
562 GlyLysGluGluGluLeuValGluLysLeuAlaGluGlyIleGluLysVa 578

1038 CTCCATCACGCGCACCACTCTCGGCCATTT..................... 1067
578 lAlaAla..AlaPheTyrProArgProValTrpTyrArgThrLeuAspAl 594

1068 .CCTAAAAAACAAACTCTTCAAGTT.........CACGACAGCCGTCAAC 1107
594 aProThrAsnGluPheArgGluMetProGlyGlyGluAspGluProGluG 611

1108 GGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCC 1157
611 luArgAsnProMetLeuGlyTrpArgGly.........IleArgArgGly 624

1158 GTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATA 1207
625 LeuAspGlnProGluLeuLeuArg...AlaGluPheLysAlaIleLysLy 640

1208 CCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTC 1257
640 sValValGluLysGlyTyrAsnAsnIleGly.................. 650

1258 GCTTTGTGCAGCTTCGTCTG.........CCCGGGCAAATACGAATACGG 1298
651 .....ValMetLeuProLeuValSerHisProGluGlnIleArgLysAla 665

1299 CCCGCTGTTGCGCAAAGTGCTGGA 1322
666 LysGluIleAlaArgSerValGly 673
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:

```
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
```

```
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;


alignment_scores:
            Quality:  112.50              Length:    536
              Ratio:  0.516                Gaps:     26
 Percent Similarity:  40.672   Percent Identity: 23.694

alignment_block:
 US-09-303-518D-131/rev x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

   1312 TGCGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGA..... 1274
        |||::::::   ||||||  |||     |||  |||
    112 CysThrSerArgGlyArgSerArgCysSerProAspArgCysCysArgTr 128

   1273 .CGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCA. 1226
           |||::::::  :::     |||     ||||||:::|||
    128 pSerArgCysSerSerProSerProArgArgCysProProSerSerProA 145

   1225 .....AAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAA 1181
             |||:::           ||||||  ::::::|||
    145 laGlyAlaProGlyAlaThrCysSerArgArgProPheSerArg..... 159

   1180 GCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATC 1131
        :::|||  ::::::  |||  |||  |||  ::::::||||||::
    160 SerArgAspSerAlaGlyProArgAlaAlaArgPheArgArgCysArgAs 176

   1130 GGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAGAG 1081
        ::::     |||     |||
    176 pAlaCysGluArgArgAlaArg..................... 183

   1080 TTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGT 1031
                                  |||              :::
    184 .........................CysProGlyProArgSerAla 190

   1030 CCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCG 981
        |||:::     :::  ||||||          ::::::     ||
    191 ProSerIleArgArgGlySerArgAspArgSerArgAlaSerArgSerAr 207

    980 ATAACGGAAATCTGAT...TGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
        |  |||  |||     ||||||::::::  :::
    207 gSerArgGlySerProLeuCysGlyAlaThrAlaThr..... 219

    933 TTGTGCAATCGCACCGTTCAATACCGAACCGGAAA..... 899
                      |||  |||:::|||:::
    220 ................SerCysProArgArgArgArgCysSerIleGly 230

    898 ...TCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGAC 852
           |||  ||||||||::  |||:::||||||     |||:::::

    231 AlaSerSerGlyCysProHisProProValArgArgSerProValAsnSe 247

    851 ACCTTCGCACCCAAAACGGTACGCA.....AGAG 823
        :  ::::::     |||  :::          ||||
    247 rSerLysArgAlaHisArgArgCysThrAlaArgArgGlyArgPheArgG 264

    822 GCGCGGTTTGTTGACTTGCAGGCCGC.....CCAAGG 791
        ||  :::       :::||||||          ||||||
    264 lyProThrSerArgAspThrGlyArgArgArgCysTrpArgTrpProArg 280

    790 CAACCACGCGCTCGG.....TATTCAGA 768
           |||:::|||          ::::::::
    281 ProArgArgCysArgCysSerArgArgTrpGlyArgProLeuTrpAlaSe 297

    767 CGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGAT 718
        :|||     |||::::::  |||     |||:::          |
    297 rGlyCysProArgAlaArgTrpArgArgGlySerAsn.....T 310

    717 GGTCCACA.....CGGTTTTATTCGCGC 695
        ||||||:::          |||      :::|||
    310 rpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysGlyArg 326

    694 CGACTG.....GCTCGATGAAATGAATGTGCGTGCCACTC 660
        |||:::          ||||||  ::::::|||  :::  ||
    327 ArgValArgSerAspThrSerAlaArgArgSerArgCysProAlaSerSe 343

    659 AAG.....CCGGCAGGAT 647
        |                |||  ||||
    343 rProIleArgTrpThrGlyArgCysArgArgTrpArgArgProLeuGlyC 360

    646 GCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGC 597
        ||:::|||     :::|||::::::          :::     ||||||
    360 ysSerProArgAlaThrCysThrAla......ArgCysGlyArgAspGly 374

    596 ACGTCTGCG......CCTGCTGCTTTACACACATGGATTTTACGTTCGGT 553
           ||||||           ||||||           ||||||::
    375 CysSerAlaPhePheGlyAsnProLeuHis.....ArgSerLe 387

    552 CAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTT 503
        :|||
    387 uArg..................................... 388

    502 TGATGATGACCGTAGGGTCGGCAGCCAGCGG.ATTGGTGTCCATCGCATT 454
                     |||      |||:::  :::  :::||||||
    389 .............GlyProTrpAlaAlaProPheArgAlaHisArgSer 400

    453 GACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACG 404
           :::      :::|||  ::::::  |||  :::      :::||
    401 ArgSerThrThrArgArgCysAlaValArgGlySerSerArgHisAspAr 417

    403 GACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGC 354
        ||||::::::  |||  |||          :::      |||:::
    417 gThrAlaSerThrArgArgProHisLysProProLysGlyCysAlaThrA 434

    353 ACTTTTTCGCTGCTCAATTTTGCCA..... 329
          :::  ::::::::::::|||
    434 spIleHisSerGlyArgTyrCysTrpProArgThrAlaSerSerArgAla 450

    328 GCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGT......TGCCT 285
        |||     :::  :::  |||||||||:::||||||     |||
    451 AlaSerGlyAlaSerAlaLysArgThrArgLeuArgArgArgSerCysPr 467

    284 TCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCAC..... 242
           |||:::  |||         :::|||||||     |||
    467 oValArgSerProArgArgArgGlyThrArgAlaAlaTrpHisSerAlaC 484

    241 ........GGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATA 200
              ::::::  |||           ::|||  :::
    484 ysGlySerSerSerArgArgProSerSerGlyArgProTrpSerValPro 500

    199 CTACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTT..... 158
        :::||||||:::|||:::|||        |||:::||||||
    501 IleArgProSerSerIleCys...GlyArgAlaValGlyLeuThrSerPr 516
```

```
     157 .....................TTTTGACGGCTTCACCTTCCTTGATTTT 130
                              |||   |||         |||      :
     516 oSerSerProLeuAsnArgProPheAlaArgArgSerAlaPro......A 531

     129 CATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGG 80
         ::|||      :::   ||||||      |||:::             |||
     531 laSerThrProCysArgArgHisAsnArgArgArgTyrGlySerArgArg 547

      79 .......................TAATGGCCGGGCCGTCATAAA..... 59
                                |||   :::::::|||
     548 ProPheArgArgArgPheAlaCysSerTrpSerSerGlnHisAspProAl 564

      58 ............................TGACTTGCTCCGGTCTGCCCGC 37
                                     |||      :::|||||||
     564 aSerGlnAspProGlnArgGlyThrCysProLeuArgAsnAlaCysProG 581

      36 GATGGGCA 29
         ||||||
     581 lyTrpAla 583
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY28654

seq_documentation_block:

```
ID   AAY28654 standard; Protein; 802 AA.
XX
AC   AAY28654;
XX
DT   01-OCT-1999  (first entry)
XX
DE   Murine Cytoplasmic phosphatase, Z70PEP protein.
XX
KW   Lymphoid Protein Tyrosine Phosphatase; Lyp protein; lymphoid cell;
KW   intracellular tyrosine phosphatase; PTPase; lymphocyte; murine;
KW   protein tyrosine kinase; PTKs; immunosuppressant; PEST sequence;
KW   T cell antigen receptor signalling; autoimmune disease; transplant;
KW   cytokine receptor signalling.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          27..288
FT                   /label= PTPase_domain
FT                   /note= "Catalytic protein tyrosine phosphatase domain"
FT   Binding-site    613..621
FT                   /label= SH3_binding_site
FT                   /note= "Proline rich sequence"
FT   Binding-site    689..695
FT                   /label= SH3_binding_site
FT                   /note= "Proline rich sequence"
FT   Binding-site    790..798
FT                   /label= SH3_binding_site
FT                   /note= "Proline rich sequence"
XX
PN   WO9936548-A1.
XX
PD   22-JUL-1999.
XX
PF   18-JAN-1999;   99WO-CA00038.
XX
PR   16-JAN-1998;   98CA-2220853.
XX
PA   (HSCR-) HSC RES & DEV LP.
XX
PI   Roifman CM;
XX
DR   WPI; 1999-444404/37.
XX
PT   New nucleic acid encoding intracellular tyrosine phosphatase and
PT   related proteins, used to modulate signalling through T cells,
PT   particularly as immunosuppressant
XX
PS   Disclosure; Page 63-64; 105pp; English.
XX
CC   The present protein sequence is that of the murine phosphatase, Z70PEP
CC   that has a single catalytic domain. The non-catalytic portion of the
CC   phosphatase contains unique sequences, including five PEST sequences
CC   rich in Pro, Glu or Asp, Ser and Thr. Z70PEP shares about 70% sequence
CC   identity with the human cytoplasmic phosphatase Lyp1. Lyp proteins are
CC   important for regulation of T cell antigen and cytokine receptor
CC   signalling and for early and late stages of T cell differentiation.
CC   Z70PEP has immunosuppressive activity. Compounds that increase
CC   expression of Lyp protein can be used as immunosuppressive agents to
CC   reduce or prevent T cell activation or proliferation, to control
CC   thymocyte differentiation, to treat autoimmune diseases and transplant
CC   situations.
XX
SQ   Sequence   802 AA;
```

alignment_scores:

Quality:	112.00	Length:	351
Ratio:	0.622	Gaps:	20
Percent Similarity:	51.282	Percent Identity:	24.501

alignment_block:

US-09-303-518D-131/rev x AAY28654 ..

Align seg 1/1 to: AAY28654 from: 1 to: 802

```
     955 GTCCCAAATAATCAT...GCGCGCCTTGTGCAATCGCACCGTTCAATACC 909
         :::   :::|||||||   :::|||   :::|||:::::   ||||||||
     297 IleSerAspAsnHisLeuGlyArgGluIleGlnAlaGlnCysSerIlePr 313

     908 GAACCGG............AAATCACGCGGTTGTCCGCGTCAACCAATTC 871
         |::::::            ::::::   :::||||||   :::     :
     313 oGluGlnSerLeuThrValGluAlaAspSerCysProLeuAspLeuProL 330

     870 GCCGG......CGGTAAGTTGAGACACCTTCG...CACCCAAAACGGTAC 830
         :::::      |||   |||:::|||   :::   |||   |||:::
     330 ysAsnAlaMetArgAspValLysThrThrAsnGlnHisSerLysGlnGly 346

     829 GCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGC 780
         ||| ||||||:::                  :::   |||:::   :::|
     347 Ala.GluAlaGlu.................SerThrGlyGlySerSerL 357

     779 TCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTC 730
         |||||:::::|||:::            |||   ::::::::::::
     357 euGlyLeuArgThrSer............ThrMetAsnAlaGluGluGlu 369

     729 TTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGA 680
         |||::::::   :::      :::           :::   |||   ||
     370 LeuValLeuHisSerAlaLysSerSerProSerPheAsnCysLeu...Gl 385

     679 AATGAATGTGCGTGCC..................ACTCAAGCCGGCAGGA 648
         |:::|||      :::                  |||:::   |||:::
     385 uLeuAsnCysGlyCysAsnAsnLysAlaValIleThrArgAsnGlyGlnA 402

     647 TGCGGGCCGCC.......................AAATTCATGTGTTTC 622
           ||||||:::                       |||::::::   :::
     402 laArgAlaSerProValValGlyGluProLeuGlnLysTyrGlnSerLeu 418

     621 GATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACA 572
         |||:::||||||::::::    |||   |||:::|||:::||||||
     419 AspPheGlySerMetLeu.PheGly...SerCysProSerAlaLeuPro. 433

     571 CATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCG...... 528
              |||      :::   |||      |||::::::   |||
     434 .....IleAsnThrAlaAspArgTyrHisAsnSerLysGlyProValLys 448

     527 CGTTTGAAGTCTTCGGCGGCTTCTTTGATG..................... 498
         |||   ||||||:::      :::|||:::
     449 ArgThrLysSerThrProPheGluLeuIleGlnGlnArgLysThrAsnAs 465

     497 .ATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCC.............. 462
          ::::::||||||::::::   |||   |||   |||
```

```
    465 pLeuAlaValGlyAspGlyPheSerCysLeuGluSerGlnLeuHisGluH 482

    461 .................ATCGCATTGACGAAGATGGCGAACGGCTCGGCA 429
                         :::   :::   :::::::|||:::   |||:::
    482 isTyrSerLeuArgGluLeuGlnValGlnArgValAlaHisValSerSer 498

    428 TCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTC..... 384
        :::      :::      |||   |||        ::::::
    499 GluGluLeuAsnTyrSerLeuProGlyAlaCysAspAlaSerCysValPr 515

    383 ....CATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCA 338
            |||:::|||      :::|||:::::         ||||||   :
    515 oArgHisSerProGlyAlaLeuArgValHisLeuTyrThrSerLeuAlaG 532

    337 ATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG 288
        ::           |||                      ||||||
    532 luAspProTyrPheSer.....................SerSerPro 540

    287 CCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTG 238
        |||:::::::|||         |||   :::      |||:::
    541 ProAsnSerAla.........AspSerLysMetSerPheAsp........ 551

    237 AATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGAT 188
                  ||||||||          |||:::||||||
    552 .............LeuProGluLysGlnAspGlyAlaThrSerProGlyA 564

    187 TCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCC 138
          :::|||   ::::::||||||      |||:::   ::::::|||
    564 laLeuLeuProAlaSerSerThrThrSerPhePheTyrSerAsnProHis 580

    137 TTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGC 88
              ::::::   :::   :::   :::::::|||   |||   |||
    581 AspSerLeuValMetAsnThrLeuThrSerPheSerProProLeuAsnGl 597

     87 GACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCC 39
           :::::::::   :::||||||      |||   :::   :::|||
    597 nGluThrAlaValGluAlaProSerArgArgThrAspAspGluIlePro 613
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220

seq_documentation_block:

```
ID   AAB94220 standard; Protein; 639 AA.
XX
AC   AAB94220;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14581.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14581; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   639 AA;
```

alignment_scores:

```
           Quality:  109.00              Length:    485
             Ratio:  0.507                 Gaps:     21
 Percent Similarity:  44.330   Percent Identity: 23.093
```

alignment_block:

US-09-303-518D-131/rev x AAB94220 ..

Align seg 1/1 to: AAB94220 from: 1 to: 639

```
   1304 AGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAG 1255
        |||:::|||::::::   ||||||   :::|||:::
     95 SerAlaProPheThrSerLeuProPheSerThrSer.............. 106

   1254 GTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCG.... 1209
         ||||||:::::::::::::::::::|||::::::::::::|||||
    107 .SerSerAlaAlaSerThrSerAsnProAsnSerAlaSerLeuSerServ 123

   1208 .....................GTATCGCCGACGATTAAATCGCGCAAA 1182
                              :::   ||||||   ::::::   :::
    123 alPheAlaGlyLeuProLeuProLeuProProThrSerGlnGlyLeuSer 139

   1181 AGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTG...CC 1135
        :::   :::      :::::::::|||   |||   |||   :::   ||
    140 AsnProThrProValIleAlaGlyGlySerThrProSerValAlaGlyPr 156

   1134 GATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCT............. 1098
        |:::|||      ::::::||||||   |||:::|||
    156 oLeuGly......ValAsnSerProLeuLeuSerAlaLeuLysGlyPheL 171

   1098 ..................................................... 1098

    171 euThrSerAsnAspThrAsnLeuIleAsnSerSerAlaLeuSerSerAla 187

   1097 ............GTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCC 1060
                    :::::::::|||:::||||||   :::::::::
    188 ValThrSerGlyLeuAlaSerLeuSerSerLeuThrLeuGlnAsn..... 202

   1059 GAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGC 1010
                                  |||      :::|||:::   |
    203 ............................SerAspSerSerAlaSerAlaP 210
```

```
1009 CGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGG 960
     |||||   :::         ||||||:::|||
 210 roAsn...LysCysTyrAlaProSerAlaIlePro..............  220

 959 TAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATAC 910
       |||:::   :::   :::|||   ::::::|||   |||
 221 ...ThrProGlnArgThrSerThrPro...GlyLeuAlaLeuProGl 235

 909 CGAACCGGAAATCACGCGGTTGTCCGCGTCAACC.............. 876
       |||:::   :::      |||:::||||||
 235 yProProSerProValAlaAsnSerThrSerThrProLeuThrLeuProV 252

 875 ..AATTCGCCG.........GCGGTAAGTTGAGACACCTTCGCACCCAAA 837
       :::||||||         |||:::|||   :::|||   ||||||
 252 alGlnSerProLeuAlaThrAlaAlaSerAlaSerThrSerAlaProVal 268

 836 ACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAAC 787
     :::       :::      :::||||||          |||   :::::
 269 SerCysGlySerSerAlaSerLeuLeuArgGlyProHisProGlyThrSe 285

 786 CACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAA 737
     :         :::         |||::::::::::      |||:::
 285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuProVal.... 300

 736 TCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGC 687
                   |||:::   |||         :::||||||
 301 .................MetIleLysThrGluProThrSerProThrPro 311

 686 TCGATGAAATGAATGTGCGTGCCACTCAAG...CCGGCAGGATGCGGGCC 640
     |||                   :::|||   |||:::      |||
 312 Ser..................AlaPheLysGlyProSerHisSerGlyAs 322

 639 GCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTG 590
      |||:::      :::      |||:::            ||||||   :
 322 nProSerHisGlyThrLeuGlyLeuSer............GlyThrLeuG 335

 589 CGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCTCAAT 540
     ::   |||         |||
 335 lyArgAlaTyrThrSerThr.............................. 341

 539 ACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGT 490
                       |||         ||||||         ::
 342 ....................SerValProIleSerLeuSerAlaCysLe 351

 489 AGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGA 440
     ::::   |||   |||||||||   |||   :::                 |
 351 uAsnProAlaLeuSerGlyLeuSerSerSerSerThrPro......LeuA 366

 439 ACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGC 390
     ||||||||         ::::::|||   |||      :::      |||
 366 snGlySerAsnProLeuSerSerIleSerLeuProProHisGlySerSer 382

 389 GCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCT 340
     :::         |||           |||           ||||||
 383 ThrProIleAlaProValPheThrAlaLeuProSerPheThrSerLeuTh 399

 339 CAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGT 290
      |||            :::|||:::         |||         ||||
 399 rAsnAsnPheProLeuThrGlyAsnProSerLeuAsnProSerValSerL 416

 289 TGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCG...... 246
     |||||::::::   |||:::|||:::   ::::::      |||
 416 euProGlySerLeuIleAlaThrSerSerThrAlaAlaThrSerThrSer 432

 245 .........CCACGGTGAATAGCGGCGATTTTG................. 222
                |||         ||||||:::|||
 433 LeuProHisProSerSerThrAlaAlaValLeuSerGlyLeuSerAlaSe 449

 221 ....CCTGAAGCCGGCGCA............................... 207
         |||   :::::::|||
 449 rAlaProValSerAlaAlaProPheProLeuAsnLeuSerThrAlaValP 466

 206 ........GTAAATACTACGCCCGGATTCTTTTTGTCTTCAAACAGCACT 165
                   ::::::|||   |||   :::   |||||||||   :::
 466 roSerLeuPheSerValThrGlnGlyProLeuSerSerSerAsnLeuSer 482

 164 TGGCCTTTTTTGACGGCTTCA......CCTTCCTTGATTTTCATCGAGGG 121
     :::|||   ::::::::::|||      ||||||:::
 483 TyrProGlyPheSerValSerAsnThrProSerValThrProAlaLeuPr 499

 120 GCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCG 71
         :::|||      ::::::||||||::::::::::|||
 499 oSerPheProGlyLeuGlnAlaProSerThrValAlaAlaValThrProL 516

  70 GGCCG 66
       |||
 516 euPro 517
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY31745

seq_documentation_block:

```
ID   AAY31745 standard; Protein; 430 AA.
XX
AC   AAY31745;
XX
DT   22-NOV-1999  (first entry)
XX
DE   Mycobacterium tuberculosis specific DNA-encoded polypeptide.
XX
KW   Tuberculosis; infection; diagnosis; DNA probe.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by TGA"
FT   Misc-difference 6
FT                   /note= "encoded by TGA"
FT   Misc-difference 20
FT                   /note= "encoded by TGA"
FT   Misc-difference 29
FT                   /note= "encoded by TGA"
FT   Misc-difference 54
FT                   /note= "encoded by TGA"
FT   Misc-difference 64
FT                   /note= "encoded by TGA"
FT   Misc-difference 69
FT                   /note= "encoded by TGA"
FT   Misc-difference 89
FT                   /note= "encoded by TGA"
FT   Misc-difference 99
FT                   /note= "encoded by TGA"
FT   Misc-difference 114
FT                   /note= "encoded by TGA"
FT   Misc-difference 119
FT                   /note= "encoded by TGA"
FT   Misc-difference 129
FT                   /note= "encoded by TGA"
FT   Misc-difference 159
FT                   /note= "encoded by TGA"
FT   Misc-difference 169
FT                   /note= "encoded by TGA"
FT   Misc-difference 182
FT                   /note= "encoded by TAG"
FT   Misc-difference 185
FT                   /note= "encoded by TGA"
FT   Misc-difference 219
FT                   /note= "encoded by TGA"
FT   Misc-difference 259
FT                   /note= "encoded by TGA"
FT   Misc-difference 269
FT                   /note= "encoded by TGA"
FT   Misc-difference 291
FT                   /note= "encoded by TGA"
FT   Misc-difference 323
```

```
FT                   /note= "encoded by TGA"
FT   Misc-difference 339
FT                   /note= "encoded by TGA"
FT   Misc-difference 349
FT                   /note= "encoded by TGA"
FT   Misc-difference 356
FT                   /note= "encoded by TGA"
FT   Misc-difference 366
FT                   /note= "encoded by TGA"
FT   Misc-difference 410
FT                   /note= "encoded by TGA"
FT   Misc-difference 421
FT                   /note= "encoded by TGA"
XX
PN   EP945462-A1.
XX
PD   29-SEP-1999.
XX
PF   25-MAR-1998;   98EP-0302287.
XX
PR   25-MAR-1998;   98EP-0302287.
XX
PA   (COUL ) CSIR COUNCIL SCI IND RES.
XX
PI   Kumar D,  Srivastava BS,  Srivastava R;
XX
DR   WPI; 1999-530042/45.
DR   N-PSDB; AAX87940.
XX
PT   New nucleic acid molecules, useful for detecting and identifying
PT   Mycobacterium tuberculosis
XX
PS   Disclosure; Page 15-27; 42pp; English.
XX
CC   The present sequence represents an amino acid sequence deduced
CC   from the Mycobacterium tuberculosis specific DNA fragment provided
CC   in AAX87940.  This DNA fragment comprises a StuI-StuI fragment of
CC   M. tuberculosis genomic DNA and contains an insertion sequence-like
CC   element and repetitive sequences.  The DNA fragment is useful as a
CC   probe, especially for detecting or identifying M. tuberculosis in
CC   clinical isolates and body fluids e.g. sputum, cerebrospinal fluid,
CC   pleural fluid, urine, gastric lavage, bronchial lavage, pericardial
CC   or lymph node aspirate (all claimed).  It is also useful for
CC   restriction fragment length polymorphism analysis of M. tuberculosis
CC   isolates (claimed).  The probe provides rapid and specific diagnosis
CC   of tuberculosis and M. tuberculosis infection.
XX
SQ   Sequence   430 AA;


alignment_scores:
            Quality:  108.50              Length:    476
              Ratio:  0.599                 Gaps:     26
 Percent Similarity:  38.025   Percent Identity: 23.109

alignment_block:
 US-09-303-518D-131 x AAY31745   ..

 Align seg 1/1  to: AAY31745  from: 1  to: 430

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||:::   :::   |||   :::|||:::      :::|||     ||
       5 Pro***SerCysCysArgArgArgSerLeuArgCysValArgGln***Pr 21

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
         |:::::::|||||||||         :::::::::
      21 oSerAlaArgCysLeuTrpGly***IleGlySer.............G 33

     132 AATCAAGGAAGGTG.....................AAGCCGTCAAAA 157
           :::|||::::::                     :::|||
      33 lyAspArgGluIleAspGlyGluSerGlyProGlyTyrArgProProPro 49

     158 AAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACT 207
         :::::   |||      :::      |||   ||| |||         |
      50 ArgSerHisCys***SerArgProCysArgLysArg.ArgTyr***ArgC 66

     208 GCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCG.............. 242
         ||         |||::::::          |||
      66 ysArgAsp***ArgSerArgCysTrpCysHisProGlyTyrSerCysArg 82

     243 ...TGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCA 289
            |||         |||:::|||            |||      |||
      83 TyrTrpCysHisProArg***SerCysArgCysTrpCysTyrArgArg** 99

     290 ACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTG 339
           |||:::
      99 *SerArgArg..................................... 102

     340 AGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGC 389
                                                       |||
     103 ..........................................Cys 103

     390 GCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGT 439
         :::|||   |||         ::::::||||||         |||
     104 CysSerCysProCysCys...ArgSerProCys***SerArgArg..... 117

     440 TCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGA.... 485
          |||            |||   :::   |||   :::||||||||||
     118 .Arg***GlyCysProCysCysSerCysGlnHis***GlyCysArgTyrC 134

     486 ........CCCTACGGTCATCATCAAAGAAGCCGC.............. 512
                 |||   :::::::::       ||||||
     134 ysArgTyrProGlySerArgTyrProSerSerArgCysProSerLeuArg 150

     513 ...CGAAGACTT.........CAAACGCGGCCTGTTGGTATTGAGCCGCC 550
            ||||||:::         :::::::|||                ||
     151 CysArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysPr 167

     551 TGACCGAACGTAAAAT......CCATGTGTGTAAAGCAGCAGG....... 587
         |:::   |||   :::         |||      |||||||||
     167 oAsn***ThrGlyArgCysCysArgCysProSerSerSerArg***GlnT 184

     588 ....................CGCAGACGTGCCGTCTGAAAATGCTGCCAA 617
                             ||||||   |||      :::||||||
     184 yr***SerProAlaGlyCysArgArgThrAlaArgCysArgCysCysCys 200

     618 TATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAG...TG 664
            |||           ||||||      |||      :::   ||
     201 CysArg.........CysTrpArgThrArgCysCysCysCysArgCysTr 214

     665 GCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGG 714
         |:::::   :::      |||         |||
     214 pGlnSerLeuGly***SerArgProArgSerArg.............. 225

     715 ACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGG 764
                 |||||||||   |||   |||         :::||||||
     226 ..........SerArgArg...CysSerArgArgArgPheGlnAsnArg 237

     765 CCGTCTGAATACCGAGCGCGTGGT...........TGCCTTGGGCGGCC 802
                   |||:::||||||            |||   :::   ||
     238 CysCys......ArgSerArgGlyPheArgIleArgCysCysSerPhePr 252

     803 TGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTG 852
         |:::   ::::::         :::         |||   |||:::
     252 oGlyPheArgAsnArgHis***IleLeuArgCysPheHisCysArgTyr* 269

     853 TCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTC 902
           |||      |||   |||                    ||||||   |||
     269 **SerCysArgArgCysArgCys.............ProArgCysPhe 280

     903 ........................................CGGTTCGG 910
                                                ||||||
     281 GlyCysArgGlyCysArgCysGlnGlyCys***SerHisArgArgPheAr 297

     911 TATTGAACGGTGC.........................GATTGCACAA 933
         ::::::|||                           :::|||:::|
```

```
 297 gCysArgGluCysCysAsnCysArgCysTrpArgCysArgGluCysSerA 314

 934 GGCGCGCATGATTATTTGGG.........................ACG 956
     |||||      |||   |||
 314 rgArgProGlyLeuProGlyArgAsp***ArgProValGlyHisArgLys 330

 957 CTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGT 1006
     :::|||        ||||||      |||   |||::::::|||   ::
 331 IleProThrCysCysPheArgCys***ArgSerProArgSerArgProAl 347

1007 TCGGCTG.........GGTTGCGCCGCAGCCGGACAAATACTCCATCACG 1047
     :              |||   ::::::::::::   |||
 347 aLeu***TrpProProGlySerCys***ThrAsnProIleArgCysCysP 364

1048 CGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGAC 1097
            |||||||||:::|||   :::      |||
 364 roSer***SerArgProIleProAlaArgProArgLeu.........Pro 377

1098 AGCCGTCAACGGCGGCGACCG......CGCCATGGTACCGATCGGCACTT 1141
     :::|||:::   |||   |||      :::   |||   :::::::::
 378 GlyArgSerTyrArgTrpProProThrLysSerGlyArgSerGlnAsnCy 394

1142 ATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGAT 1191
           |||::::::
 394 sTrpHisArgSerSer.............................G 400

1192 TTAATCGTCGGCGATACCGACAGCGC 1217
       :::|||   |||||||||:::|||
 400 lySerArgThrArgTyrArgHisArg 408

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04954

seq_documentation_block:
ID   AAY04954 standard; Protein; 572 AA.
XX
AC   AAY04954;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 41T#2.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34206.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 41T; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   572 AA;


alignment_scores:
            Quality:  108.50              Length:    507
              Ratio:  0.580                Gaps:     31
 Percent Similarity:  36.884   Percent Identity: 21.105

alignment_block:
 US-09-303-518D-131 x AAY04954   ..

 Align seg 1/1  to: AAY04954  from: 1  to: 572

  81 CGAAGTCGCGTTGCTTGGCGA.................AGAATATGTCG 112
     |||      ||||||||||||                 |||   |||
 157 ArgArgTrpValAlaTrpArgCysArgGluLeuSerSerArgCysCysGl 173

 113 GCATGCGCCCCTCGATGAAAATCAAGGA..................... 140
           |||      ::::::::::::
 173 yProProProGluProSerThrSerAlaArgThrArgGlyTrpProValA 190

 141 .......................AGGTGAAGCCGTCAAAAAAGGCCAAGT 167
                               |||   ||||||   |||||||||
 190 rgSerProProProValThrArgArgSerSerArgTrpLysArgProPro 206

 168 GCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTT 217
                                   :::|||:::   |||:::   ::
 207 ........................AsnSerLeuThrCysSerProMe 214

 218 CAGGCAAAATCGCCGCTATTCACCG.......................T 243
     :|||         |||:::::::|||                       |
 214 tArg......AlaArgTrpThrProAlaAspGlyAlaCysTrpPheCysT 229

 244 GG...CGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAA 290
     ||   |||   |||      |||:::::::|||            |||:::
 229 rpThrArgCysAlaProProSerAlaGlyArgHisLeuProGlyArgSer 245

 291 .........CGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
             ||||||   |||   |||      |||
 246 ThrAsnProArgArgAlaArgCysArgProThrArgLeu........... 258

 332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
                        :::||||||   |||   :::   |||
 259 ................ProAsnAlaPro...ProArgAsnSerArg***C 269

 382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAA................... 413
         |||      :::      :::|||:::
 269 ysIleCysTrpArgTyrValMetLeuGlnArgArgThrSerCysGlyIle 285

 414 ...........................AATCCCTGCCGTAGATGCCG 433
                                  :::|||      |||  |||
 286 AspSerArgAsnTrpValSerArgTrpProSerProLeuLeuArg..Pro 301

 434 AGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATC...CGCTGGCT 480
     :::   :::|||   :::|||          |||||||::       |||
 302 ThrAlaThrProTyrThrSerThrProThrThrProValProProTrpLy 318

 481 GCCGACCCTACGG.....................TCATCATCAAAGAAGC 509
      |||      |||                     :::|||   ::::::
 318 sProAspTrpArgTrpGlyGluLeuAlaGlySer***SerArgArgServ 335

 510 CGCCGAAGACTTCAAAC.........................GCGGCC 532
       |||      ::::::                         |||
 335 alProGlyProAlaAspCysArgProValAlaGlyArgGlyAlaAlaPro 351

 533 TGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCA 582
     ||||||      ::::::      :::|||          |||
 352 CysTrpArg...SerSerThrAlaThrValProProSerCysSerProGl 367

 583 GCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGA 632
```

```
     :::|||   :::|||
367 yArgAlaProAlaCys.............................. 372

633 ATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCA 682
                                                  :::
373 ..........................................CysAsp 374

683 TCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGAC 732
        |||:::         :::|||
375 ArgValGlnThrPro***HisArgPro..................... 383

733 GTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCG 782
                                    :::   :::|||::::
384 ..............................ProIleSerValProThrS 390

783 CGTGGTTG...CCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGC 829
    ::|||      ||||||::::::      :::   :::   :::   |||
390 erTrpCysGlyProTrp***Thr...ProAlaProArgThr***TrpCys 405

830 GTACCG......TTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAA 873
      |||       :::|||                    ||||||   ||
406 CysProMetAlaMetTrp.....................ProProLysAs 415

874 TTGG.........TTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATT 914
    ||||         :::   |||   :::|||         :::   |||
415 nTrpTrpProGlyValProArgArgSerAlaGlyAlaSerThrTrpTyr. 431

915 GAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACA 964
                          :::|||         |||
432 ................ProCysArgProAspArgTrpCysArgGlyTrp 442

965 ATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCT.. 1012
      |||:::|||       :::   |||      |||   :::|||
443 ProArgTrpProCysMetThrArgProAlaArgProSerThrThrAlaTh 459

1013 ............................GGGTTGCGCCGC.......... 1024
                                   |||     |||
459 rAlaTrpProValProProValLeuProGlyThrAspArgCysAlaLeuP 476

1025 ..AGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCC......... 1063
        :::   :::::::|||:::   :::|||   |||
476 roProLysArgArg***ProGlyProValProAlaSerArgAlaThrVal 492

1064 ...ATTTCCTAAAAAACAAACTCTTCAAGTTCA..............CG 1095
       :::|||      |||:::   :::||||||               ||
493 TrpValSerArgAlaThrArgCys***SerSerProThrMetSerProAr 509

1096 ACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGA 1145
     |:::||||||:::   :::|||:::   |||:::|||:::|||:::
509 gArgProSerValTrpSerThrCysCysTrpHisArgGluAlaIle.... 524

1146 GCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAA 1195
                |||:::   |||
525 .............Trp***ArgCys....................... 528

1196 TCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTT...GCTTGGAATTG 1242
       :::      |||::::::   :::   |||:::   :::||||||
529 ...***LeuAlaProAla***ProLysThrTrpLeuSerSerTrpAsnGl 544

1243 GACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGA 1292
               |||::::::   |||   |||   :::   :::      :
544 yMetCysThrThrThrIleGlnAlaProSerTrpSerProThrAlaProA 561

1293 ATACGGCC......CGCTGT 1306
     ::||||||      ||||||
561 spThrAlaAlaThrArgCys 567
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198

seq_documentation_block:
ID ABB64198 standard; Protein; 2406 AA.
XX
AC ABB64198;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 19386.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08301.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 19386; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2406 AA;

alignment_scores:
Quality: 108.00 Length: 473
Ratio: 0.519 Gaps: 28
Percent Similarity: 43.975 Percent Identity: 22.833

alignment_block:
US-09-303-518D-131/rev x ABB64198 ..

Align seg 1/1 to: ABB64198 from: 1 to: 2406

```
1298 CCGTAT.................TCGTATTTGCCCGGGCAGACGAAGCT 1267
     ||||||                 ||||||||||||           ::
1266 ProTyrSerGlyLeuThrHisGlySerTyrLeuProProValLeuProVa 1282

1266 G.................................CACAAAGCGAGGT 1253
     :                                 |||::::::   |
1282 lAlaThrProAsnLeuSerAsnLeuProProThrGlnHisArgSer...S 1298

1252 CTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCG 1203
     ||:::|||   ||||||::::::      |||
1298 erAspSerArgAsnSerArgGluSerProAla..................  1308

1202 CCGACGATTAAATCGCGCAAAAGCAAGGTAGGC...AGGATGTCCAACGG 1156
        ::::::||||||      |||:::::|||   :::::::|||   ::
1309 ...SerLeuLysSerThrProSerAsnIleGlyLeuAsnValSerMetAl 1324
```

```
1155 CATTACG...CGCTCATAAGTGCCGATCGGT.............ACCA 1124
1324 aProThrLeuArgSerIleThrProLeuAsnAsnSerSerAlaIleSerS 1341

1123 TGGCGCGGTCGCCGCCGTTGACGGCTGTCGTG.........AACTTGAAG 1083
1341 erGlyAlaSerGlnProValValSerValValProSerAlaAsnSerThr 1357

1082 AGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTG.........CGCGTGAT 1042
1358 AlaLeuSerMetSerAsn...ProHisIleSerHisSerHisHisValPr 1373

1041 GGAGTATTTGTCCGGC.....................TGCGGCGCAACCC 1013
1373 oAlaTyrAlaSerGlyAlaPheSerSerSerAlaAlaAlaGlyThrSerT 1390

1012 AGCCGAACAGCTCTTTG......CTGCGGCCTTCTTCGATAACGGAAATC 969
1390 hrProAsnSerGlyLeuSerThrLeuAlaValThrSerLeuSerThr... 1405

 968 TGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACC 919
1406 .......................SerAlaAlaPro.........GlnPr 1411

 918 GTTCAATACCGAACCGGAAATCACGCGG...TTGTCCGCGTCAACCAATT 872
1411 oHisSerHisPheProGlnSerThrGlnMetLeuProGlnSerGlyAsnP 1428

 871 CGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGG 822
1428 heSerServalSer..................................... 1432

 821 CGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATT 772
1433 ...HisLeuLeuThrThrHisProMetSerSerGlnAsnGlnProMetVa 1448

 771 CAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAAT 722
1448 lArg.................................................. 1449

 721 TGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATG 672
1450 .....CysGlySerThrLeuTyrSerGlnSerSerAla............. 1460

 671 TGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTT.. 624
1461 .................AlaAlaThrAlaProProSerAlaAlaAlaAl 1471

 623 .....................................TCGATATTGGCAG 611
1471 aValSerAsnPheThrProSerValLeuAlaValGlnSerLeuThrThrA 1488

 610 CATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTA 561
1488 laValThrSerSerSerSerSerProSerThrLeuSerSerServalIle 1504

 560 CGTTCGGTC.................AGGCGGCTCAATACCAACAGGCC 529
1505 GlnLysValIleSerProLysGlyGluSerProCysAsnLysAspArgAs 1521

 528 GCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAG 479
1521 pSerSerTyrSerSerProAlaAsnAlaValValThrThrCysAlaProT 1538

 478 CCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCA 429
1538 hrThrProIleValSer.....................SerGlySerAla 1547

 428 TCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAA 379
1548 ArgProThrProProLeuSerAsn.......................... 1555

 378 GCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCA 329

1556 .....................CysThrSerMetGlyIleGlyMetValA 1565

 328 GCGCTTCAGGTACGTAGCGTTCG......AACTCGATTTCG......TCG 291
1565 snAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSer 1581

 290 TTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACG 241
1582 IleProAlaThr..................................... 1585

 240 GTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCG 191
1586 .......AlaGlyIleHisValSerAla............ThrAsnProS 1596

 190 GATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCT 141
1596 erPhe......GlnSerSerSerTyrPheProThrProLeuAlaProPro 1610

 140 TCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAA 91
1611 Pro..............................SerSerProSerPr 1616

  90 CGCGACTTCGGTAATGGCC 72
1616 oAlaThrSerSerAlaAla 1622
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

```
ID   AAB59827 standard; Protein; 1592 AA.
XX
AC   AAB59827;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #4 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
```

CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1592 AA;

alignment_scores:
Quality: 109.50 Length: 550
Ratio: 0.517 Gaps: 26
Percent Similarity: 38.545 Percent Identity: 22.727

alignment_block:
US-09-303-518D-131/rev x AAB59827 ..

Align seg 1/1 to: AAB59827 from: 1 to: 1592

```
1312 TGCGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGA..... 1274
     |||::::::   ||||||  |||     |||  |||
 707 CysThrSerArgGlyArgSerArgCysSerProAspArgCysCysArgTr 723

1273 .CGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCA. 1226
           |||::::::   :::     |||     ||||||:::|||
 723 pSerArgCysSerSerProSerProArgArgCysProProSerSerProA 740

1225 .....AAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAA 1181
           |||:::           ||||||   :::::::|||
 740 laGlyAlaProGlyAlaThrCysSerArgArgProPheSerArg..... 754

1180 GCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATC 1131
     :::|||   ::::::  |||  |||  |||  :::::|||||::
 755 SerArgAspSerAlaGlyProArgAlaAlaArgPheArgArgCysArgAs 771

1130 GGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAGAG 1081
     ::::     |||     |||
 771 pAlaCysGluArgArgAlaArg..................... 778

1080 TTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGT 1031
                                                  |||
 779 ..............................................Cys 779

1030 CCG.....................................GC 1026
     |||                                   ::
 780 ProGlyProArgSerAlaProSerIleArgArgGlySerArgAspArgSe 796

1025 TGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAAC 976
     :   |||:::   ||||||  ::      ||||||
 796 rArgAlaSerArgSerArgGlySerProLeuCysGly.......... 808

 975 GGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAA 926

 809 ..............................................Ala 809

 925 TCGCACCGTTCAATACCGAACCGGAAA.................TCACG 894
     :::      |||   |||:::|||:::                 |||
 810 ThrAlaThrSerCysProArgArgArgArgCysSerIleGlyAlaSerSe 826

 893 CGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGC 844
      |||||||||:::   |||:::||||||      |||::::::   ::::
 826 rGlyCysProHisProProValArgArgSerProValAsnSerSerLysA 843

 843 ACCCAAAACGGTACGCA.....................AGAGGCGCGGTT 815
     ::      |||   :::                    ||||||   :::
 843 rgAlaHisArgArgCysThrAlaArgArgGlyArgPheArgGlyProThr 859

 814 TGTTGACTTGCAGGCCGC.................CCAAGGCAACCACG 783
               :::|||||||                ||||||      ||
 860 SerArgAspThrGlyArgArgArgCysTrpArgTrpProArgProArgAr 876

 782 CGCTCGG.........................TATTCAGACGGCCTGT 760
     |:::|||                        ::::::::::|||
 876 gCysArgCysSerArgArgTrpGlyArgProLeuTrpAlaSerGlyCysP 893

 759 TACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACA 710
       |||::::::   |||      |||:::            ||||||:::
```

```
 893 roArgAlaArgTrpArgArgGlySerAsn............TrpSerSer 905

 709 ........................CGGTTTTATTCGCGCCGACTG.. 689
                              |||      :::|||||:::
 906 GlyArgSerSerAlaAlaSerProLysArgThrCysGlyArgArgValAr 922

 688 ............GCTCGATGAAATGAATGTGCGTGCCACTCAAG..... 657
                 ||||||   :::::::|||   :::   |||
 922 gSerAspThrSerAlaArgArgSerArgCysProAlaSerSerProIleA 939

 656 ............................CCGGCAGGATGCGGGCCG 639
                                 |||   ||||||:::|||
 939 rgTrpThrGlyArgCysArgArgTrpArgArgProLeuGlyCysSerPro 955

 638 CCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGC 589
         :::|||::::::        :::      ||||||   |||||
 956 ArgAlaThrCysThrAla......ArgCysGlyArgAspGlyCysSerAl 970

 588 G......CCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGC 545
     |              ||||||           ||||||:::|||
 970 aPhePheGlyAsnProLeuHis...........ArgSerLeuArg.... 981

 544 TCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATG 495

 981 .................................................. 981

 494 ACCGTAGGGTCGGCAGCCAGCGG.ATTGGTGTCCATCGCATTGACGAAGA 446
         |||      |||:::   :::  :::||||||      :::
 982 ......GlyProTrpAlaAlaProPheArgAlaHisArgSerArgSerTh 996

 445 TGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTG 396
     :::|||   ::::::  |||  :::      :::||||||:::
 996 rThrArgArgCysAlaValArgGlySerSerArgHisAspArgThrAlaS 1013

 395 CGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTC 346
     ::   |||  |||            :::      |||:::   :::
1013 erThrArgArgProHisLysProProLysGlyCysAlaThrAspIleHis 1029

 345 GCTGCTCAATTTTGCCA.....................GCGCTTCA 321
     :::::::::::::|||                      |||
1030 SerGlyArgTyrCysTrpProArgThrAlaSerSerArgAlaAlaSerGl 1046

 320 GGTACGTAGCGTTCGAACTCGATTTCGTCGT......TGCCTTCAACGGC 277
      :::  :::  |||||||||:::||||||      |||      |||:
1046 yAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArgS 1063

 276 AATCACGACTGACTGAAGTACGCGCTTTTCGCCAC.............. 242
     ::   |||        :::||||||     |||
1063 erProArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySer 1079

 241 GGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCC 192
     ::::::   |||           :::|||   :::      :::||||||
1080 SerSerArgArgProSerSerGlyArgProTrpSerValProIleArgPr 1096

 191 GGATTCTTTTTGTCTTCAAACAGCACTTGGCCTT.............. 158
     |:::|||:::|||         |||:::||||||
1096 oSerSerIleCys...GlyArgAlaValGlyLeuThrSerProSerSerP 1112

 157 ............TTTTGACGGCTTCACCTTCCTTGATTTTCATCGAGG 122
                 |||   |||        |||      :::|||
1112 roLeuAsnArgProPheAlaArgArgSerAlaPro......AlaSerThr 1126

 121 GGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGG........ 80
        :::   ||||||      |||:::           |||
1127 ProCysArgArgHisAsnArgArgArgTyrGlySerArgArgProPheAr 1143

  79 ..............TAATGGCCGGGCCGTCATAAA............. 59
                     |||   :::::::|||
1143 gArgArgPheAlaCysSerTrpSerSerGlnHisAspProAlaSerGlnA 1160

  58 ..................TGACTTGCTCCGGTCTGCCCGCGATGGGCA 29
                         |||      :::||||||   ||||||
1160 spProGlnArgGlyThrCysProLeuArgAsnAlaCysProGlyTrpAla 1176
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242

seq_documentation_block:
ID   AAB27242 standard; Protein; 571 AA.
XX
AC   AAB27242;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human EXMAD-20 SEQ ID NO: 20.
XX
KW   Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW   inflammation; reproductive disorder; cardiovascular disorder;
KW   immune disorder; musculoskeletal disorder; developmental disorder;
KW   gastrointestinal disorder; cell proliferation disorder.
XX
OS   Homo sapiens.
XX
PN   WO200068380-A2.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US12811.
XX
PR   11-MAY-1999;   99US-0133643.
PR   23-AUG-1999;   99US-0150409.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Bandman O,  Hillman JL,  Tang YT,  Lal P,  Yue H,  Baughn MR,  Lu DAM;
PI   Azimzai Y;
XX
DR   WPI; 2001-007395/01.
DR   N-PSDB; AAC66909.
XX
PT   Isolated polynucleotide encoding extracellular matrix or
PT   adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT   preventing disorders associated with expression of EXMAD such as
PT   proliferative, immune and genetic disorders -
XX
PS   Claim 1; Page 106-107; 129pp; English.
XX
CC   The present invention provides the protein and coding sequences for 25
CC   novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC   These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC   EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC   EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC   EXMAD-20, EXMAD-21,  EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC   useful in the prevention and treatment of cancers, cell proliferation,
CC   cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC   gastrointestinal disorders and inflammation.
XX
SQ   Sequence   571 AA;


alignment_scores:
           Quality:  106.50            Length:    514
             Ratio:  0.467               Gaps:     19
 Percent Similarity:  44.358   Percent Identity: 18.872

alignment_block:
 US-09-303-518D-131/rev x AAB27242   ..

 Align seg 1/1  to: AAB27242  from: 1  to: 571

    1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTC 1291
         |||   :::      :::|||:::|||   :::   |||:::   :::::
      23 ProMetMetProThrThrSerGlyThrSerGlnAlaSerSerSerPheAs 39

    1290 GTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCA 1241
         :           :::|||:::||||||         |||   :::|||:
      39 nThrAlaLysThrSerThrSerLeuHis.....SerHisThrSerS 53

    1240 ATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTA..... 1206
         :::::   :::|||::::::         |||:::
      53 erThrHisHisProGluValThrProThrSerIleThrAsnIleThrLeu 69

    1205 .....TCGCCGACGATTAAATCGCGCAAAAG 1180
                        :::|||:::      :::   :::::
      70 AsnProThrSerIleGlyThrTrpThrProValAlaHisThrThrSerAl 86

    1179 CAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATCG 1130
         :      :::|||:::::::::   :::|||          |||   |
      86 aThrSerSerArgLeuThrThrProPheThrThrHisSerProProThrG 103

    1129 GTACC......ATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTG 1086
         ||:::      ::::::   :::   |||:::||||||:::
     103 lySerSerProIleSerSerThrGlyProMetThrAlaThrSerPheGln 119

    1085 AAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTA 1036
            :::   ::::::
     120 ThrThrThrTyrTyrThrPro..... 126

    1035 TTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTT 986
                          ::::::|||:::::::::|||
     127 ...............ProSerHisProGlnThrThrLeu..... 134

     985 CTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCG 936

     134 ..... 134

     935 CCTTGTGCAATCGCACCGTTCAATACCGAA.....CCGGAAATC.. 897
         |||      :::   ||||||:::|||:::         |||:::
     135 ProThrHisValProProPheSerThrSerLeuValThrProSerThrHi 151

     896 .....ACGCGGTTGTCCGCGTCAACCAATTCGC 869
                          |||:::::::::::|||::::::
     151 sThrValIleIleThrThrHisThrGlnMetAlaThrSerAlaSerIleH 168

     868 CGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGC 819
           ::::::         |||      |||   |||:::  |||
     168 isSerThrProThrGlyThrValProProProThrThrLeuLysAlaThr 184

     818 GGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAG 769
         |||                |||||   ::::::|||
     185 GlySerThrHisThrAlaProProMetThrValThr..... 196

     768 ACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGA 719
                                              ||||||      :
     197 ..............................ThrSerGlyThrSerG 202

     718 TGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGC 669
         :::::|||:::   :::   :::   |||:::|||   :::   :::
     202 lnThrHisSerSerPheSerThrAlaThrAlaSerSerSerPheIleSer 218

     668 .....GTGCCACTCAAGCCGGCAGGATGCGGGCC 640
                          :::|||   :::   ::::::      ||
     219 SerSerSerTrpSerSerTrpLeuPro...GlnAsnSerSerSerArgPr 234

     639 GCCAAATTCATGTGTTTCG..... 621
         ||||:::|||   ::::::
     234 oProSerSerProIleThrThrGlnLeuProHisLeuSerSerAlaThrT 251

     620 .....ATATTGGCAGCATTTTCAGACGGCACG 594
                          :::   ::::::||||||   :::
     251 hrProValSerThrThrAsnGlnLeuSerSerSerPheSerProSerPro 267

     593 TCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCT 544
         |||||||||:::::::::   :::::::::      |||
     268 SerAlaProSerThrValSerSerTyrValProSerSerHisSerSerPr 284

     543 CAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGA 494
          :::|||:::   |||   :::   :::|||::::::   :::
     284 oGlnThrSerSerProSerValGlyThrSerSerSerPheValSerAlaP 301
```

```
    493 CCGTAGGGTCGGCA......GCCAGCGGATTGGTGTCCATCGCATTGACG 450
          |||   |||:::      :::||||||      |||        |||
    301 roValHisSerThrThrLeuSerSerGlySerHisSerSerLeuSerThr 317

    449 AAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACG 400
            ::::::::::|||:::|||::::::   ::::::   ::::::
    318 HisProThrThrAlaSerValSerAlaSerProLeuProSerSerPr 334

    399 GGTGCGAAGCGCAGTCCATAAG.........CCTGATTGAATCAGGTTGC 359
         :::   |||::::::   :::         |||      |||
    334 oAlaAlaSerThrThrIleArgAlaThrLeuProHisThrIleSerSerP 351

    358 GGCGCACTTTTTCGCTGCTC.........AATTTTGCCAGCGCTTCAGGT 318
            |||:::|||   |||          :::   :::::::::|||
    351 roPheThrLeuSerAlaLeuLeuProIleSerThrValThrValSerPro 367

    317 ACGTAGCGTTCGAACTCGATTTCGTCG.........TTGCCTTCAACGGC 277
        |||      |||:::      ||||||          :::||||||||||
    368 ThrProSerSerHisLeuAlaSerSerThrIleAlaProSerThrPr 384

    276 AATCACGACTGACTGAAGTACGCGCTTTTCGCCA.............. 243
            ||||||      ||||||:::   :::|||
    384 oArgThrThrAla...SerThrHisThrAlaProAlaPheSerSerGlnS 400

    242 ................CGGTGAATAGCGGCGATTTTGCCTGAAGCCGGC 210
                                  :::::::::   :::|||   :::|||
    400 erThrThrSerArgSerThrSerLeuThrThrArgValProThrSerGly 416

    209 GCAGTAAATACTACGCCCGGATTC.....................TTTTT 181
          |||:::   |||   |||
    417 PheValSerLeuThrSerGlyValThrGlyIleProThrSerProValTh 433

    180 GTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTT 131
         :::   ::::::      |||      |||   |||||||:::      |
    433 rAsnLeuThrThrArgHisProGlyProThrLeuSerProThrThrArgP 450

    130 TCATCGAGGGGCGCATGCCGACATAT...TCTTCGCCAAGCAACGCGACT 84
        ||:::   :::   :::   ::::::   |||:::|||:::::|||
    450 heLeuThrSerSerLeuThrAlaHisGlySerThrProAlaSerAlaPro 466

     83 TCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTG 42
                  |||             |||      |||:::
    467 ValSerSerLeuGlyThrProThrProThrSerProGlyVal 480
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG93052

seq_documentation_block:

ID AAG93052 standard; Protein; 720 AA.
XX
AC AAG93052;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6806.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH68271.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 6806; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 720 AA;

alignment_scores:

Quality:	106.50	Length:	508
Ratio:	0.447	Gaps:	31
Percent Similarity:	46.850	Percent Identity:	23.228

alignment_block:
US-09-303-518D-131 x AAG93052 ..

Align seg 1/1 to: AAG93052 from: 1 to: 720

```
     46 CCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC...GCGTT 92
        ||||||:::::::::      :::|||   ::::::::::::   |||
     58 ProGluGluLeuVal......SerProGlnValSerGlnIleTyrAlaSe 72

     93 GCTTGGCGAAGAATATGTC.....................GGCATGCGCC 121
          ||||||      :::                      |||   |||
     72 rAspGlyGluThrGluLeuAlaArgIleValProProGluGlyAsnArgG 89

    122 CCTCGATGAAAATCAAGGAA...GGTGAAGCCGTCAAAAAAGGCCAAGTG 168
          :::   |||   :::      :::::||||||::::::   |||
     89 lnMetValThrIleAspGlnValProAspThrValLysAsnAlaValVal 105

    169 CTGTTTGAAGACAAAAAG.........AATCCGGGCGTAGTATTTACTGC 209
          ||||||::::::         |||||||||      :::|||::
    106 AlaAlaGluAspArgGluPheTyrThrAsnProGlyPheSerIleThrGl 122

    210 GCCGGCTTCAGGCAAAATCGCCGCTATT..............CACCGTG 244
        :   |||   :::   :::::::::|||                |
    122 yTyrAlaArgAlaAlaLeuGlyValIleThrGlyAspSerSerAlaGlyG 139

    245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGAC 294
        ||   :::   :::   |||:::   |||      :::      :::|||
    139 lyGlySerThrIleThrGlnGlnTyrValLysLysAlaValValGlyAsp 155

    295 GAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTG......GCAAAATT 338
        |||   ::::::   |||   :::   |||   :::      ||||||::
    156 GluArgSerLeuIleArgLysAlaLysGluLeuValTyrSerAlaLysMe 172

    339 GAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACT. 387
        :::::::|||      :::   ::::::::::::::   |||   |||
    172 tAlaAsnGluTrpSerLysAspGluValLeuGluAlaTyrLeuAsnThrV 189

    388 ........................................GCGCTT 393
                                                |||:::
```

```
 189 alTyrPheGlyArgAsnAlaTyrGlyValGlnAlaAlaAlaHisAlaPhe 205

 394 CGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGC 443
       :::|||   :::::::::   |||::::::::::      :::||
 206 PheAspLysProValGluGluLeuThrAlaAlaGluGlyAlaValLeuAl 222

 444 CATCTTCGTC.........AATGCGATGGAC......ACCAATCCGCTGG 478
       |     :::         :::   :::|||      |||||||||:::
 222 aAlaSerIleGlnLeuProSerGlnLeuAspProTrpThrAsnProValG 239

 479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGC 528
       |||:::                                     |||
 239 luAlaGlu.....................................ThrArg 243

 529 GGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAA 578
           |||::::::   |||:::|||
 244 TrpAsnTyrValMetAspGlyLeuValGlu..................... 253

 579 AGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATC......G 622
           ||||||        :::|||:::|||:::           |
 254 ....IleGlyAlaIleSerAlaGluGluArgAlaValAlaThrTyrProG 269

 623 AAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCAC 672
     |||||   :::               ||||||:::   |||:::
 269 luThrThrAsp..............ProAlaSerAsnSerAla...... 278

 673 ATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAA 722
         :::   |||   :::|||:::|||                |||::
 279 ......TyrThrGluAlaThrGlyThrAsn.........GlyLeuIleLy 290

 723 TTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGA 772
     :   |||   :::      ::::::::::|||   :::
 290 sAsnGlnValMetAlaGluLeuSerGluLeuGlyIle.............. 302

 773 ATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAAC......... 813
       ||||||      |||          ||||||||||||::
 303 ..ThrGluAspAspValGlnThrArgGlyLeuGlnValThrThrThrIle 318

 814 AAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTAC 863
       |||:::      :::   ::::::   |||      :::||||||.
 319 AspProLysThrGlnGluGlyAlaValGluAlaValGlnAsnGlnLeu.. 334

 864 CGCCGGCGAATTGGTTGACGCGGACAACCGC......GTGATTTCCGGTT 907
           :::|||:::::   :::||||||      |||:::|||   :
 335 .......AspLeuLeuSerGluAsnAsnArgAlaAlaValValSerIleA 349

 908 CGGTATTGAACGGTGCGATTGCA......CAAGGCGCGCATGATTATTTG 951
     ::      ||||||||||:::           |||:::::|||
 349 spProSerAsnGlyAlaValArgAlaTyrTyrGlyGlyGluAspAlaThr 365

 952 GGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGA 1001
     |||   :::            :::   ::::::   |||   :::   ::
 366 Gly...TrpAspPheAlaAsnAlaProLeuGlnThrGlySerThrPheLy 381

1002 GCTGTTCGGCTGGGTTGCG.................CCGCAGCCGGACA 1033
     ::::||||||   :::|||                  |||      :::
 381 sIlePheGlyLeuAlaAlaAlaLeuGlnGlnGlyIleProLeuSerGlnP 398

1034 AATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTC 1083
       ||||||   :::   :::|||:::
 398 roTyrSerSerAlaProValThrVal...................... 406

1084 TTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGAT 1133
                              ||||||      :::      ::
 407 .........................GlyAspAlaGlnIleGlyAsnVa 414

1134 CGGC.....................ACTTATGAGCGCGTAATGCCGTTGG 1162
     :|||                     :::   |||:::::::::   |||
 414 lGlyGlySerGlyCysGlySerCysSerIleGluGlnAlaLeu...LeuH 430

1163 ACATCCTGCCTACCTTGCTTTTGCGC.........GATTTAATCGTCGGC 1203
               |||   ::::::|||         ||||||      |||
 430 isSerTyrAsnThrSerPheIleArgLeuGlnGlnAspLeuGluAsnGly 446

1204 ......GATACC...GACAGCGCGCAGGCTTTGGGTTGCTTGGAA..... 1239
           ||||||   |||   |||:::|||||||||      :::
 447 SerGlnAspThrAlaAspMetAlaHisAlaLeuGlyIleAlaLysSerLe 463

1240 ..................TTGGACGAAGAAGACCTCGCTTTGTGCAGCT 1270
                         |||   |||:::      :::      :::
 463 uProThrIleProGluThrLeuThrGluAsnGlyGluThrProTyrGluG 480

1271 TCGTCTGCCCGGGCAAATACGAATACGGCCCGCTG...TTGCGCAAAGTG 1317
       :::      |||:::||||||      ||||||   :::   ::::::
 480 lyIleIleLeuGlyGlnTyrGluSerArgProLeuAspMetAlaSerAla 496

1318 CTGGAAACCATTGAGAAGGAAGGC 1341
     :::   ||||||   :::||||||
 497 MetAlaThrIleAlaAsnGluGly 504
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21954

seq_documentation_block:

ID ABG21954 standard; Protein; 1078 AA.
XX
AC ABG21954;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21945.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86141.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52313; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human

```
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1078 AA;


alignment_scores:
            Quality:  106.50              Length:    486
              Ratio:  0.461               Gaps:       21
 Percent Similarity:  47.531   Percent Identity: 21.811

alignment_block:
 US-09-303-518D-131/rev x ABG21954    ..

 Align seq 1/1  to: ABG21954  from: 1  to: 1078

   1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGC..............  1302
        :::|||:::::::|||:::|||:::|||   :::|||
    320 AlaPheAlaLeuValThrSerSerLeuSerGluSerValPheLeuArgAs  336

   1301 .......GGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAG  1259
                |||      :::::::|||            :::|||   |||:
    336 pValPheGlyLeuSerAlaPheLeu.........ArgSerLeuAlaLysS  350

   1258 CGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCG  1209
        :::::|||   ::::::   |||         ::::::          |||
    350 erAsnSerLeuGlyLysIleSerPheLeuSerSerSerLeuLeuAlaSer  366

   1208 GTATCGCCGACGATTAAATCG.....................CGCAAAAG  1180
        :::|||   ||||||:::|||                     ||||||::
    367 AlaSerValThrIleSerSerProSerThrSerSerLeuGlnArgLysAl  383

   1179 CAAGGTAGGCAGGATG...TCCAACGGCATTACGCGCTCATAAGTGCCGA  1133
        ::::   :::|||:::   |||:::|||:::                |||
    383 aSerPheAlaArgLeuProSerSerGlyLeu..............ProS  395

   1132 TCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAG  1083
           |||      |||          :::   |||   :::||||||
    395 erGluThrCysIleArgLeuAlaSerAlaSerProValLysGlyLeuLys  411

   1082 AGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTT  1033
        :::                          |||   :::   ::::::||
    412 ArgVal.....................SerValLeuIleThrSerPheLe  421

   1032 GTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTT  983
        ||||   :::|||::::::   |||:::   :::      :::|||||||
    421 uSerLeuAlaGlySerSerProProArgIleAlaSerSerGluProSerS  438

    982 CGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCT  933
        ||:::   :::                              |||
    438 erLeuPheAsp................................AlaVal  443

    932 TGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTGTCCGC  883
           ::::::::::|||   :::::::::             :::::::||
    444 SerSerMetGlyProValSerSerArg............SerMetAspAl  456

    882 GTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGG  833
        ||||::::::          :::|||
    456 aSerSerSerLeuSerProThrSer.......................... 464

    832 TACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACG  783
                             ||||||:::   ||||||:::|||:::
    465 ....................LeuThrAlaThrProProSerAlaSerPro  474

    782 CGCTCG.........GTATTCAGACGGCCTGTTACGAACAAACGTCCGAT  742
           |||          :::      |||   :::|||:::   |||
    475 SerSerAspValAlaAlaSerLeuArgAlaAlaThrSer***ArgCysSe  491

    741 AGCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGA  692
           :::   :::                    ::::::   ||||||   |
    491 rLysPheArgAla.................AlaAlaSerPheAlaSerT  502

    691 CTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGG  642
        ||:::      |||            |||   :::|||:::   ::::::
    502 hrAlaPheSerLys............ProCysSerProSerGluAlaAla  514

    641 ...CCGCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCAC  595
                ::::|||   |||||||||   |||:::||||||   :::::
    515 LeuSerValSerSerThrValSerIleLysAlaCysPheSerGlyAlaSe  531

    594 GTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGC  545
        :|||                          :::||||||   :::      :
    531 rSer.......................LeuLeuArgProThrAlaSerP  540

    544 TCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATG  495
        ::   :::      |||      ::::::      :::      ::::::
    540 heAlaAlaAlaCysProMetGlySerAspProProSerProSerLeuLeu  556

    494 ACCGTAGGGTCGGCAGCCAGC..........................GG  472
        |||   :::::::::|||:::                          ||
    557 ThrProSerAlaSerAlaAlaProSerAlaProLeuSerMetProGlyGl  573

    471 ATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTACGG  422
        |:::      ::::::::::::::   ||||||:::::::::|||::::
    573 yIleSerLeuPheSerValAsnSerSerAlaAsnSerGlyCysSerAlaV  590

    421 CAGGGATT...TTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT  375
        ::||||::      |||:::|||         ::::||      :::|||
    590 alGlyPheHisAlaLeuLysGlyAlaSer***ThrAlaProAlaSerPro  606

    374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC  325
                                  |||                ||||||
    607 His......SerAlaAlaSerProPheProCysThrMetSerAlaSerLe  621

    324 TTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGCAA  275
                        |||:::||||||:::|||:::   |||::::::
    621 uCysTrpGlyAlaSerSerSerSerIleAlaSerPhePheSerSerValT  638

    274 TCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGATT  225
          :::                  |||      :::         ::::::
    638 yrVal.................PheCysThrAsnCysTrpAspSerLeu  648

    224 TTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTG..... 180
        :::         :::||||||:::::|||      |||::::::
    649 PheGluLeuIleAlaAlaValSerSerThrMetLysPheIleValGlyAl  665

    179 ................TCTTCAAACAGCACTTGGCCTTTTTTGACGGCTT  146
                        ::::::|||::::::      ||||||
    665 aProArgAlaArgThr***AlaAsnGlyAsnGluSerPheLeuProLeuV  682

    145 CACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCT......  102
           :::      |||:::   |||      |||:::   |||
    682 alSerThrSerSerPheLeuGlyGlyPheProProSerSerSerLeuLeu  698

    101 TCGCCAAGCAACGCGACT............................TC  82
        :::|||:::::|||:::                              ||
    699 ThrProAspSerAlaSerGlnTrpGlnTrp***ProAsnAlaLysGlySe  715

     81 GGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCCGCGATGG  32
        ||||:::      |||:::      :::      |||   ||||||:::
    715 rValIleIleCysProAlaProSerSerAsnProGlyArgProAlaPheL  732

     31 GCAGATTT 24
           |||
    732 euSerPhe 734

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504

seq_documentation_block:
ID   AAW87504 standard; Protein; 1061 AA.
XX
AC   AAW87504;
XX
DT   23-FEB-1999  (first entry)
```

```
XX
DE   Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW   NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS   Homo sapiens.
XX
PN   US5849895-A.
XX
PD   15-DEC-1998.
XX
PF   20-APR-1994;   94US-0231193.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI   Daggett LP,  Lu C;
XX
DR   WPI; 1999-069812/06.
DR   N-PSDB; AAV82910.
XX
PT   DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT   assembly of functional glutamate receptor subunits
XX
PS   Example 3; Columns 261-268; 203pp; English.
XX
CC   The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC   receptor subunit (NMDAR). The nucleic acid sequence does not contain the
CC   860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891)
CC   between nucleotides 1300 and 1301, an additional 24 nucleotides
CC   (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in
CC   AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR
CC   subunits contribute to the formation of NMDA-activated cation-selective
CC   ion channels. In addition to being useful for the production of NMDA
CC   receptor subunit proteins, the nucleic acids are also useful as probes to
CC   identify and isolate nucleic acids encoding related receptor subunits.
CC   Functional glutamate receptors can be assembled from several NMDA
CC   receptor subunit proteins of one type (homomeric) or from combinations of
CC   subunit proteins of different types (heteromeric). The present invention
CC   also comprises methods for using such receptor subunits to identify and
CC   characterise compounds which affect the function of such receptors , e.g.
CC   agonists, antagonists and modulators of glutamate receptor function. The
CC   invention also comprises methods for determining whether unknown
CC   protein(s) are functional as NMDA receptor subunits.
XX
SQ   Sequence   1061 AA;


alignment_scores:
            Quality:  106.00              Length:    490
              Ratio:  0.613                Gaps:     28
 Percent Similarity:  35.306   Percent Identity: 23.265

alignment_block:
 US-09-303-518D-131 x AAW87504   ..

 Align seg 1/1  to: AAW87504  from: 1  to: 1061

      26 ATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCC 75
         ::::::|||||||||         :::   :::           |||||
     689 LeuAlaProSerArgIleGlyValAlaAlaAlaValArgProHisArgPr 705

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC 125
         |      ::   |||   ||||||          ||||||   |||
     705 oProAlaArgProArgGlyLeuAlaPro...AlaHisAlaCysProPro. 720

     126 GATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGG...CCAAGTGCTGT 172
                    ::::::||| |||:::|||   |||   |||
     721 .............ProThrArgPro.GlnSerArgAlaProArgAlaGl 732

     173 TTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGC 222
             ||||||         :::|||   :::   |||||||||:::|||
     732 yAspArgGlnThrGlyValAlaArgArgLeu...CysAlaGlyLeuArg. 747

     223 AAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGT...ACTTCAGTCAGT 269
         :::   |||   :::|||   |||::::::|||          :::|||
     748 ..SerProArgAlaAlaProArgArgArgGlyArgProCysProThrSer 763

     270 CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTAC 319
         :::||||||      |||      :::|||
     764 ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyPr 780

     320 CTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
         :::||||||      ::::::      |||:::|||
     780 oGlyThrAlaGlyGlyThrSerArgProProSerGlyPro..... 793

     370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCC 419

     793 ..................................................  793

     420 TGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCA 469

     793 ..................................................  793

     470 ATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC 519
                 |||||||||                          |||   :
     794 .........CysArgPro.......................ArgAlaV 799

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         ::   |||   |||      :::||||||   |||   :::      |||
     799 alThrThrAlaProPhe...LeuGluProThrAspProAla...AlaPro 813

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGC...TGCCA 616
                 |||||||||      |||:::         |||   |||::
     814 SerSerArgSerSerArgSerProArgSerTrpArgThrCysArgCysSe 830

     617 ATATCGAAACACATGAATTTGG.........CGGCCCGCATCC....... 650
         :   |||::::::      |||         ||||||:::::
     830 rValArgSerSer......TrpProGlyGlyArgProCysThrArgProG 845

     651 ..............................TGCCGGCTTGAGTGGCAC 668
                                      |||       |||
     845 lyProGlyAlaArgAlaArgValThrLeuProCysProAlaProTrp... 860

     669 GCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCA 718

     860 ..................................................  860

     719 TCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGT 768
                                                     |||||||
     861 .....................................ProArgProS 864

     769 CTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACC 818
         ||         ||||||      ||||||   ||||||
     864 erLeuGlyProAlaArgCysProLeuGlyAlaProAlaProProAlaPro 880

     819 GCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCG 868
         ||||||   |||   |||                ::::::::::::||||||
     881 AlaProThrAlaThrArg.............ProAlaGlyAlaTrpArgAr 893

     869 GCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAAC 918
         ||||                                          :::|
     893 gArg.....................................SerA 896

     919 GGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCA 968
         |||||   |||
     896 rgCysAlaCys..................................... 899

     969 GATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGT.. 1016
             |||         ||||||   :::||||||   |||   |||
     900 ......ArgSerThrGlyArgProAlaArgArgAlaSerArgGlnGlyPr 914

    1017 .....................................................T 1017
```

```
                                                                      |
    914 oProProGlySerThrAspSerThrSerAlaCysThrProThrProThrC 931

   1018 GCGCCGCAGCCGGACAAATACTCC.................ATCACGCG 1049
        ||   :::||||||   ::: |||                  :::|||
    931 ysHisCysAlaGlyGlyLeu.SerValLeuThrPheHisProValThrAl 947

   1050 CACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTC.............. 1086
         |||:::   |||            :::   |||:::
    947 aThrAlaProGlySerProAlaProGlyGlyLeuTrpGlyThrAlaAlaG 964

   1087 .................AAGTTCACGACAGCCGTCAACGGCGGCGACCGC 1119
                        :::   |||   :::|||:::       |||
    964 lyLeuTrpGlyTrpAlaGlnAlaThrGluThrValGlyAspTrpThrArg 980

   1120 GCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCT 1169
        :::   ||||||:::|||
    981 SerAlaValProValGlyArgLysAlaSer..................... 990

   1170 GCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGC 1219
                       ||||||         ||||||
    991 .................ArgAspProAlaProGlyAsp............. 997

   1220 AGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC 1269
              ||| :::      |||:::::::|||   :::|||   |||::
    998 ........Gly.SerProValTrpSerGlnLysCysGluLeuSerAlaTh 1011

   1270 TTCGTCTGCCCGGGCAAATACGAATACGGCCCGCT.........GTTGCG 1310
        ::::::: |||:::|||:::         ||||||         |||
   1011 rGlnAla.ProSerGlnLeuAspSerLeuProAlaThrValArgValLys 1027

   1311 CAAAGTGCTGGA 1322
        ::::::||||||
   1028 ArgGlnAlaGly 1031
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503

seq_documentation_block:

```
ID   AAW87503 standard; Protein; 1212 AA.
XX
AC   AAW87503;
XX
DT   23-FEB-1999  (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW   NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS   Homo sapiens.
XX
PN   US5849895-A.
XX
PD   15-DEC-1998.
XX
PF   20-APR-1994;   94US-0231193.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI   Daggett LP,  Lu C;
XX
DR   WPI; 1999-069812/06.
DR   N-PSDB; AAV82909.
XX
PT   DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT   assembly of functional glutamate receptor subunits
XX
PS   Example 3; Columns 253-262; 203pp; English.
XX
CC   The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC   receptor subunit (NMDAR). The nucleic acid sequence does not contain
CC   the 366 5'-most nucleotides, by the insertion of 11 nucleotides between
CC   nucleotides 1300 and 1301, nor the 15 nucleotides at positions
CC   1960-1974, nor the 1061 3' nucleotides, as set forth in AAV82889. The
CC   cDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute
CC   to the formation of NMDA-activated cation-selective ion channels. In
CC   addition to being useful for the production of NMDA receptor subunit
CC   proteins, the nucleic acids are also useful as probes to identify and
CC   isolate nucleic acids encoding related receptor subunits. Functional
CC   glutamate receptors can be assembled from several NMDA receptor subunit
CC   proteins of one type (homomeric) or from combinations of subunit proteins
CC   of different types (heteromeric). The present invention also comprises
CC   methods for using such receptor subunits to identify and characterise
CC   compounds which affect the function of such receptors , e.g. agonists,
CC   antagonists and modulators of glutamate receptor function. The invention
CC   also comprises methods for determining whether unknown protein(s) are
CC   functional as NMDA receptor subunits.
XX
SQ   Sequence   1212 AA;
```

alignment_scores:

```
            Quality:  106.00              Length:    490
              Ratio:  0.613                 Gaps:     28
 Percent Similarity:  35.306   Percent Identity: 23.265
```

alignment_block:

US-09-303-518D-131 x AAW87503 ..

Align seg 1/1 to: AAW87503 from: 1 to: 1212

```
     26 ATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCC 75
        ::::::|||||||||       :::  :::          |||||
    840 LeuAlaProSerArgIleGlyValAlaAlaAlaValArgProHisArgPr 856

     76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC 125
        |      :::   |||   ||||||           ||||||   |||
    856 oProAlaArgProArgGlyLeuAlaPro...AlaHisAlaCysProPro. 871

    126 GATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGG...CCAAGTGCTGT 172
                       ::::::||| |||:::|||   |||   |||
    872 .............ProThrArgPro.GlnSerArgAlaProArgAlaGl 883

    173 TTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGC 222
           ||||||         :::|||   :::   ||||||||||:::|||
    883 yAspArgGlnThrGlyValAlaArgArgLeu...CysAlaGlyLeuArg. 898

    223 AAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGT...ACTTCAGTCAGT 269
          :::   |||   :::|||   |||:::::::|||          :::|||
    899 ..SerProArgAlaAlaProArgArgArgGlyArgProCysProThrSer 914

    270 CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTAC 319
           :::|||||||       |||      :::|||
    915 ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyPr 931

    320 CTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
           :::|||||||      ::::::      |||:::|||
    931 oGlyThrAlaGlyGlyThrSerArgProProSerGlyPro.......... 944

    370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCC 419

    944 .................................................. 944

    420 TGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCA 469

    944 .................................................. 944

    470 ATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC 519
                  |||||||||                        |||   :
    945 ..........CysArgPro.......................ArgAlaV 950

    520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
        ::   |||   |||      :::|||||||   |||   :::      |||
```

```
 950 alThrThrAlaProPhe...LeuGluProThrAspProAla...AlaPro 964

 570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGC...TGCCA 616
           |||||||||      |||:::       |||   |||::
 965 SerSerArgSerSerArgSerProArgSerTrpArgThrCysArgCysSe 981

 617 ATATCGAAACACATGAATTTGG.........CGGCCCGCATCC..... 650
     :   |||::::::      |||         ||||||::::::
 981 rValArgSerSer......TrpProGlyGlyArgProCysThrArgProG 996

 651 ..............................TGCCGGCTTGAGTGGCAC 668
                                   |||         |||
 996 lyProGlyAlaArgAlaArgValThrLeuProCysProAlaProTrp... 1011

 669 GCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCA 718

1011 ................................................. 1011

 719 TCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGT 768
                                                 |||||||
1012 .....................................ProArgProS 1015

 769 CTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACC 818
     ||        ||||||      ||||||   ||||||
1015 erLeuGlyProAlaArgCysProLeuGlyAlaProAlaProProAlaPro 1031

 819 GCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCG 868
     ||||||   |||   |||              :::::::::::|||||
1032 AlaProThrAlaThrArg...........ProAlaGlyAlaTrpArgAr 1044

 869 GCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAAC 918
     ||||                                        :::|
1044 gArg.....................................SerA 1047

 919 GGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCA 968
     |||||   |||
1047 rgCysAlaCys..................................... 1050

 969 GATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGT.. 1016
         |||         ||||||   :::||||||   |||   |||
1051 ......ArgSerThrGlyArgProAlaArgArgAlaSerArgGlnGlyPr 1065

1017 ..............................................T 1017
                                                    |
1065 oProProGlySerThrAspSerThrSerAlaCysThrProThrProThrC 1082

1018 GCGCCGCAGCCGGACAAATACTCC.................ATCACGCG 1049
     ||   :::||||||   ::: |||                :::|||
1082 ysHisCysAlaGlyGlyLeu.SerValLeuThrPheHisProValThrAl 1098

1050 CACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTC............. 1086
      |||:::   |||          :::   |||:::
1098 aThrAlaProGlySerProAlaProGlyGlyLeuTrpGlyThrAlaAlaG 1115

1087 ................AAGTTCACGACAGCCGTCAACGGCGGCGACCGC 1119
                     :::   |||   :::|||:::         |||
1115 lyLeuTrpGlyTrpAlaGlnAlaThrGluThrValGlyAspTrpThrArg 1131

1120 GCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCT 1169
     :::   ||||||:::|||
1132 SerAlaValProValGlyArgLysAlaSer..................... 1141

1170 GCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGC 1219
                    ||||||         ||||||
1142 ................ArgAspProAlaProGlyAsp............. 1148

1220 AGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC 1269
           ||| :::      |||:::::::|||   :::|||   |||::
1149 ........Gly.SerProValTrpSerGlnLysCysGluLeuSerAlaTh 1162

1270 TTCGTCTGCCCGGGCAAATACGAATACGGCCCGCT.........GTTGCG 1310
     ::::::: |||:::|||:::          ||||||         |||
1162 rGlnAla.ProSerGlnLeuAspSerLeuProAlaThrValArgValLys 1178
```

```
1311 CAAAGTGCTGGA 1322
     ::::::|||||
1179 ArgGlnAlaGly 1182
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB60656

seq_documentation_block:

ID ABB60656 standard; Protein; 2088 AA.
XX
AC ABB60656;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 8760.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04759.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 8760; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2088 AA;

alignment_scores:

Quality:	106.00	Length:	538
Ratio:	0.453	Gaps:	19
Percent Similarity:	43.494	Percent Identity:	20.260

alignment_block:
US-09-303-518D-131/rev x ABB60656 ..

Align seg 1/1 to: ABB60656 from: 1 to: 2088

```
1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
     :::||||||   :::   :::::::::          :::|||:::   ::
1297 LysProSerValGluThrThrThrThrProAlaProThrSerAlaGlnPh 1313

1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
```

```
     :|||:::      ||||||::::::      |||      :::::::|||:
1313 eSerPheGlyPheGlyGlnSerAsnGlnGlyLysAspValAlaAspSerL 1330

1243 CCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATT 1194
     ::::::::::::   ||||||:::       ::::::||||||
1330 ysLysThrGluAlaProLysThrPheMetPheGlyValSer..... 1343

1193 AAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTC 1144
     |||   ::::::   |||:::      :::::::::||||||   :::::
1344 LysValGluGluProLysThrValSerPheGlyThrGlyIleLysGluTh 1360

1143 ATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCG 1094
     :   :::      :::|||   |||   :::|||      :::|||::::
1360 rThrAlaThrSerSerThrGluAlaThrAlaProThrProAlaAlaAlaA 1377

1093 TGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTG 1044
     ::               |||      |||   |||   ::::::   :::
1377 laProAlaProValGlnPheValPheLysAlaProThrThrAlaThrThr 1393

1043 ATGGAGTATTTGTCC............................... 1029
        :::   |||:::
1394 AlaSerSerLeuThrThrThrIleSerThrThrSerAsnAlaProAlaLe 1410

1028 ....GGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTT 983
         |||   :::         |||:::||||||            |||
1410 uGlyGlyPheSerPheGlyAlaProSerSerSerSerThrValSerSerS 1427

 982 CGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCT 933
     ||   |||:::                                |||:::   |||
1427 erThrThrSerThr.....................SerAlaAsnPro 1435

 932 TGTGCAATCGCACCGTTC.............................. 915
     :::|||:::   |||:::
1436 AlaAlaValLysProMetPheSerTrpSerGlyAlaGlySerAlaValSe 1452

 914 .............AATACCGAACCG........................ 903
                 :::   :::|||
1452 rSerThrSerSerSerGlnGlnProValAlaLysAlaProThrLeuGlyP 1469

 902 ................GAAATCACGCGGTTGTCCGCGTCAACCAAT... 873
                        :::|||      ::::::||||||:::
1469 heGlyValSerSerSerThrValThrThrThrThrThrSerThrLysVal 1485

 872 .........TCGCCGGCGGTAAGTTGAGACACCTTCGCA.......... 843
             :::||||||   :::   |||      |||
1486 PheAlaPheThrProAlaSerGlyLeuAspProAlaAlaAlaThrSerAl 1502

 842 .........................................CCCAAAA 836
                                                 |||   |
1502 aProAlaAlaGlyAlaGlyPheSerPheGlySerGlnSerLysProAlaT 1519

 835 CGGTACGC.....................AAGAGGCGCGGTTTGTTG 810
     ||::::::                        :::      ::::::
1519 hrThrGlnAsnThrGlyThrPhePhePheGlyGlnProThrAlaValAla 1535

 809 ACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGT 760
        :::   |||   :::   ::::::   |||:::|||      |||::
1536 ProAlaThrProThrAsnProSerValSerSerIlePheGlyAlaProAl 1552

 759 TACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACA 710
     :|||:::            :::   :::|||            :::   |
1552 aThrSerThrThrAlaSerThrSerValSerAlaThrThrSerThrSerT 1569

 709 CGGTT..............TTATTCGCGCCGACTGGCTCGATGAAATGA 675
     ||:::               ||||||||||||:::::
1569 hrAlaAsnAlaIleAlaSerSerPheAlaProThrSerThr......... 1582

 674 ATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGT 625
            |||        ::::::   |||      ::::::   ::
1583 .........ProGlnLeuPheGlyAsnTrpGlyGluLysLysThrAspLe 1596

 624 TTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTAC 575
     ::::   ::::::|||   |||   :::   ::::::   :::

1596 uThrThrPheGlyAlaSerSerGlySerGlyThrThrThrThrProSerP 1613

 574 ACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGT 525
        |||                          ::::::|||
1613 heGlyTrp.......................SerSerAsnGlyAspAla 1621

 524 TTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAG 475
       ||||||::::::|||:::            ||||||||||||||||
1622 AlaLysSerAsnSerAlaAla............ValGlySerAlaAla.. 1633

 474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTA 425
                                    :::   ::::::||||||||||
1634 ............................ValProSerSerSerAlaSerT 1641

 424 CGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT 375
     ||   :::      :::   |||
1641 hrMetAlaThrProIlePheGly........................... 1648

 374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC 325
                           :::   |||:::::::::   :::|||::
1649 ....................SerSerSerMetPheGlyProSerSerSe 1658

 324 TTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCT...TCAACGG 278
     :|||::::::      |||:::|||   :::|||||||||   ::::::|
1658 rSerAsnAsnThrThrSerThrSerThrThrSerLeuProPheGlySerA 1675

 277 CAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCG 228
     ||   ||||||                                  ||||||
1675 laAlaThrThr................................AlaAla 1680

 227 ATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTGTC 178
           |||   :::|||   :::      |||      |||
1681 ThrThrProAlaGlyGlyAsnAlaAlaLeuThrGlyLeuPheGlyAsnVa 1697

 177 TTCAAACAGC..........................ACTTGGCCTTTTT 155
      :::||||||                         |||   |||
1697 lGlyAsnSerLeuAlaGlyValGlyAlaProValAlaThrThrProAlaA 1714

 154 TGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGC............ 117
       ||||||:::      |||      |||   |||:::
1714 laThrAlaAlaAlaProLeuThrAsnIlePheGlyAsnProThrProVal 1730

 116 .........ATGCCGACATATTCTTCG............CCAAGCAACGC 88
                |||::::::::::|||            ||||||   ::
1731 AlaAlaAlaAlaProValPheGlySerGlySerThrIleProSerAlaGl 1747

  87 GACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCCG 38
     :   ::::::   |||:::   :::   :::::::::   :::||||||:
1747 yPheGlyAlaProAlaAlaAlaAlaProLeuAlaAlaProAlaLeuProG 1764

  37 CGATGGGCAGATTT 24
     ::      :::|||
1764 lyAlaPheAsnPhe 1768
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48589

seq_documentation_block:

```
ID   AAU48589 standard; Protein; 372 AA.
XX
AC   AAU48589;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #9485.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
```

```
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59543.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 9784; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   372 AA;


alignment_scores:
            Quality:  105.50              Length:    374
              Ratio:  0.659                Gaps:     18
 Percent Similarity:  42.781   Percent Identity: 21.658

alignment_block:
 US-09-303-518D-131/rev x AAU48589   ..

 Align seg 1/1  to: AAU48589  from: 1  to: 372

    1280 GGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCA 1231
         |||   ::::::   ::::::::::|||:::|||   |||:::|||   ::
      21 GlyThrSerSerThrArgArgSerArgGlnSerArgSerGlySerProSe 37

    1230 A.............................................C 1229
         :                                             |
      37 rTrpProSerSerValThrThrProValAlaArgArgThrCysLeuThrP 54

    1228 CCAAAGCCTGCGCGCTGTCG............GTATCGCCGACGATTAAA 1191
         ||      |||:::   |||              ||||||   ::::::
      54 roProLeuCysGlyThrSerProSerProProProSerProArgLeuSer 70

    1190 TCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATA 1141
         ||||||   :::   :::      ::::::::::::::::::   :::
      71 SerArgLeuAlaThrThrProValLeuThrThrAlaLeuAlaSerThrVa 87

    1140 AGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGA 1091
              :::   |||   :::   :::|||||||      :::
      87 lProThrValArgThrSerSerSerLysProProGlyProSerCys.... 102

    1090 ACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATG 1041
                                  ||||||
     103 .........................TrpProPro.............. 105

    1040 GAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCG 991
              ::::::   ::::::      |||:::::::|||   :::
     106 .........AlaAlaProSerSerLeuThrProThrThrSerAlaIleSe 119

     990 GCCTTCTTCG..................................... 981
         ||||||
     119 rThrSerSerAlaProProSerProLeuArgCysThrGlySerArgCysA 136

     980 ........ATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGC 939
                 :::|||::::::      |||   |||            :::
     136 rgCysTrpLeuThrArgLeuLeuArgTrpThrArgAla......ProAla 150

     938 GCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTT 889
         :::|||||||||:::   |||      |||:::|||   :::
     151 SerProCysAlaVal...ProSerValThrSerProThrLeuLeuGly.. 165

     888 GTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCA 839
          :::::::::   |||   |||                      ||||||:
     166 .GlyGlyAsnCysAsnCysProPro.................AlaProS 176

     838 AAACGGTA.....................CGCAAGAGGCGCGGTTTGTTG 810
         :::::|||                    |||  |||   |||
     176 erSerValValThrAspGlySerSerProArgProArgSerGlySerSer 192

     809 ACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGT 760
         :::      |||||||||   |||   :::|||      |||   |||
     193 AlaProAspProProLysAsnThrArgGlnSerArgAlaArgProPro.. 208

     759 TACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACA 710
             :::   |||::::::   :::
     209 .......SerProProValSerSerSer..................... 215

     709 CGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTC 660
                                                        |||
     216 .............................................Leu 216

     659 AAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGC 610
         ||||||:::      :::|||   :::      ||||||
     217 LysProSer...TrpSerProAlaArgTrpThrValSer........... 228

     609 ATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTAC 560
                       |||   |||:::|||      |||          |
     229 .............ProSerLeuProSerAlaLysProThrSerThrArgA 241

     559 GTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCG 510
         ||:::::::   |||   :::::::::   |||         :::||||||
     241 rgAlaThrSerArgTrpArgSerSerAlaProGlySerGlyThrSerAla 257

     509 GCTTCTTTGATGATGACCGTA..............GGGTCGGCAGCC.. 477
         :::      :::|||||||:::                  |||||||||
     258 ThrValGlyValMetThrIleCysAlaThrProCysTrpSerAlaAlaGl 274

     476 .......AGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCT 434
                |||||   :::|||   :::::::|||   |||   :::||||
     274 yAsnLeuSerGlySerLeuSerThrCysValThrValMetArgThrGlyS 291

     433 CGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTC 384
         ||            |||   :::   :::|||   :::    ::::::::
     291 er.........ArgGlySerMetAlaThrGlySerLeu.AlaValSerGl 304

     383 CATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTT 334
         :   |||         |||   ||||||      :::      :::
     304 ySerSerAlaSerProSerGlnCysGlyIleArgTrpAlaProThrAlaS 321

     333 TGCCAGCGCTTCAGGTACGT 314
```

```
          |||:::::::::::::|||
     321 erProThrMetSerThrArg 327

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU39645

seq_documentation_block:
ID   AAU39645 standard; Protein; 338 AA.
XX
AC   AAU39645;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #541.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59508.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 840; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   338 AA;


alignment_scores:
            Quality:  105.00            Length:    379
              Ratio:  0.695               Gaps:     18
 Percent Similarity:  39.842   Percent Identity: 22.427
```

```
alignment_block:
 US-09-303-518D-131/rev x AAU39645   ..

 Align seg 1/1  to: AAU39645  from: 1  to: 338

   1031 TCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTC 982
        ::::||||||::::::            ||||||      |||   |||
     17 ThrGlyCysAlaGlyLeuThrSerLeuSerSerThrThrArgAlaSerHi 33

    981 GATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTT 932
                                                 ||||||:::|||
     33 sSer..............................SerCysValProV 39

    931 GTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTGTCCGCG 882
                                 |||:::|||      :::   ||||||:::
     39 al.................ThrSerProGlySerSerSerLeuSerSer 49

    881 TCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGT 832
        |||::::::   |||::::::        :::   |||::::::
     50 ProThrThrThrSerAlaIleAlaArgArgArgLeuArgProArgSerPr 66

    831 ACGCAAGAGGCGCGGTTTGTTGACTTGC..................... 804
           :::            |||||||||
     66 oThrGln............LeuThrCysSerGlyTrpProArgArgSerI 79

    803 .............AGGCCGCCCAAGGCAACCACGCGC............ 780
                         |||   :::   |||||||||
     79 leProAlaAlaProAlaProThrSerProThrThrArgAlaArgAsnThr 95

    779 .........TCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGC 739
                 |||:::   ||||||      |||:::      ||||||
     96 ProProGlnSerAlaThrArgArgPheProThrSerLeuTrpProIleAr 112

    738 AATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTG 689
        ||||||
    112 gThrThrSer..................................... 115

    688 GCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGA......... 648
                         |||:::                 |||
    116 .............ArgCysLeuAlaThrAlaArgProGlyTyrSerIle 127

    647 .......................TGCGGGCCGCCAAATTCATGTGTTTC 622
                                 ||||||   ::::::||||||::
    128 SerThrThrSerSerTrpCysArgValGlyProSerSerAlaCysValTh 144

    621 GATATTGGCA................................GCATTTT 605
        :   |||:::                                |||::::
    144 rAlaLeuSerAlaGlyIleThrValThrThrArgSerGluCysAlaLeuT 161

    604 CAGACGGCACGTCTGCGCCTGCTGCTTTA...CACACATGGATTTTACGT 558
        ::   :::||||||   |||   |||         ||||||
    161 hrProSerThrSerProProProAlaThrGlyPheThrTrp......... 174

    557 TCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGC 508
                                                       ::
    175 ..............................................Se 175

    507 TTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCG 458
        ::::   |||:::|||:::::::::::::::::   :::|||:::::::|||
    175 rThrThrMetLeuThrThrSerGlyThrSerArgAlaLeuThrAlaIleA 192

    457 CATTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTG 408
                  :::         |||||||:::          ::::::
    192 rgArgLeuTyrIle.........SerAlaAlaArgTyrProValValGly 205

    407 AACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGAT.............. 372
        :::   |||                 :::   |||:::
    206 SerArgArgPheProArgProTyrArgThrProSerCysGlyAlaMetPr 222

    371 .TGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGCTT 323
           :::|||:::   |||         :::   :::::::::|||:::
    222 oThrLeuArgIleSerArgProThrTrpMetProSerMetThrSerCysI 239
```

```
     322 CAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCA...ACGGCA 276
     239 leGlyThrThrArgArgGly......SerProIleProSerLeuThrAla 253

     275 ATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATA........ 234
     254 SerThrIleValArgSerSerArgArgSerValArgTrpValArgSerSe 270

     233 ....GCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGAT 188
     270 rSerSerSerProHisProArgThrSerProThrGlyThrThrIleGlyS 287

     187 TC......TTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCA 144
     287 erThrArgPheArgProSerAlaArgSerTrp.................. 297

     143 CCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAG 94
     298 .....SerSerProSe 301

      93 CAACGCGACTTCG............GTAATGGCCGGG 69
     301 rSerAlaThrThrSerArgAsnGlyValMetThrGly 313
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44890

seq_documentation_block:

```
ID   AAR44890 standard; Protein; 532 AA.
XX
AC   AAR44890;
XX
DT   22-JUN-1994  (first entry)
XX
DE   Diphtheria toxin delta-142-147-148 mutant.
XX
KW   DT; protein exotoxin; NAD-dependent ADP-ribosyltransferase; vaccine;
KW   diphtheria toxoid; deletion mutant; mutein; variant; double mutant;
KW   reversion mutation; site-directed mutagenesis.
XX
OS   Corynebacterium diphtheriae.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..532
FT                   /note= "Diphtheria toxin mutant; Glu(142), Val(147)
FT                          and Glu(148) have been deleted"
XX
PN   WO9325210-A.
XX
PD   23-DEC-1993.
XX
PF   17-MAY-1993;   93WO-US04606.
XX
PR   18-JUN-1992;   92US-0901712.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Collier RJ,  Killeen K,  Mekalanos J;
XX
DR   WPI; 1994-007178/01.
DR   N-PSDB; AAQ54338.
XX
PT   New DNA encoding diphtheria toxin deletion mutants - with no
PT   toxicity and low risk of reversion, and derived toxoids and
PT   transformed cells, useful in vaccines
XX
PS   Claim 11; ; 42pp; English.
XX
CC   Oligonucleotide-directed mutagenesis of the wild-type diphtheria
CC   gene (specifically the region encoding the DT-A fragment) results
CC   in deletion of the codons for Val-147 and active site residue
CC   Glu-148. The resulting mutein is not toxic, making it useful in
CC   diphtheria vaccines. The risk of reversion to toxicity is much
CC   lower for the 147-148 double mutant than for the prior art 148
CC   single mutant, while its immunogenicity is not impaired. The
CC   147-148 mutein opt. has other amino acid residues substd. or
CC   deleted, e.g. wild-type Glu(142) is deleted. The specification
CC   includes the wild-type DT amino acid sequence (see AAR44888) but does
CC   not include any mutant sequences; the wild-type sequence was modified
CC   according to the description in the claims to give AAR44890.
XX
SQ   Sequence   532 AA;
```

alignment_scores:

```
            Quality:  105.00            Length:    461
              Ratio:  0.536               Gaps:     25
 Percent Similarity:  42.516   Percent Identity: 20.390
```

alignment_block:

US-09-303-518D-131 x AAR44890 ..

Align seg 1/1 to: AAR44890 from: 1 to: 532

```
      31 CCCATCGCGGGCAGACCGGAGCAAGTCATT...TATGACGGCCCGGCCAT 77
      72 ProLeuSerGlyLysAlaGlyGlyValLysValThrTyrProGlyLe 88

      78 TACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGA 127
      88 uThrLysValLeu.............................AlaL 94

     128 TGAAAATCAAGGAAGGTGAAGCCGTCAAAAAA..................... 159
      94 euLysValAspAsnAlaGluThrIleLysLysGluLeuGlyLeuSerLeu 110

     160 .........................GGCCAAGTGCTGTTTGAAGACAAAAA 185
     111 ThrGluProLeuMetGluGlnValGlyThrGluGluPheIleLysArgPh 127

     186 GAATCCGGGC.........GTAGTATTTACTGCGCCGGCTTCAGGCAAAA 226
     127 eGlyAspGlyAlaSerArgValValLeuSerLeuProPheAlaGlySerS 144

     227 TCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATT 276
     144 erSerTyrIleAsnAsnTrpGluGlnAlaLysAlaLeuSerValGluLeu 160

     277 GCCGTTGAAGGCAACGACGAAATCGAGTTCGAA..................... 309
     161 ..................GluIleAsnPheGluThrArgGlyLysArgGl 171

     310 ................CGCTACGTACCTGAAGCGCTGGCAAAATTGAGCA 343
     171 yGlnAspAlaMetTyrGluTyrMetAlaGlnAlaCysAla.........G 185

     344 GCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA............. 381
     185 lyAsnArgValArgArgSer...ValGlySerSerLeuSerCysIleAsn 200

     382 ......TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGT 425
     201 LeuAspTrpAspValIleArgAspLysThrLysThrLysIleGluSerLe 217

     426 AGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGC 475
     217 u......LysGluHisGlyProIleLysAsnLysMetSerGluSerProA 232

     476 TGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
     232 snLysThrValSerGluGluLysAlaLysGlnTyrLeuGluGluPheHis 248

     526 CGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGTGTG 575
     249 GlnThrAlaLeuGluHisProGluLeuSerGluLeuLysThr........ 262
```

```
    576 TAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAA 625

    262 ..................................................... 262

    626 CACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATT 675
                                        ::::::||||||:::
    263 ..............................ValThrGlyThrAsnPro 268

    676 CATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTA 725
          |||          :::||||||||||   :::::::|||:::::|||
    269 ValPhe.........AlaGlyAlaAsnTyrAlaAlaTrpAlaValAsnVa 282

    726 TCAAGACGTGATT..................................... 738
            :::||||||
    282 lAlaGlnValIleAspSerGluThrAlaAspAsnLeuGluLysThrThrA 299

    739 ....................GCTATCGGACGTTTG......TTCGTAACA 762
                             :::||||||   :::      ::::::
    299 laAlaLeuSerIleLeuProGlyIleGlySerValMetGlyIleAlaAsp 315

    763 GGCCGTCTG......AATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCA 806
        |||   :::      |||||||||::::::||||||
    316 GlyAlaValHisHisAsnThrGluGluIleValAla.............. 327

    807 AGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTC 856
              :::         |||   :::::::::   |||::::::
    328 .......GlnSerIleAlaLeuSerSerLeuMetValAlaGlnAlaIleP 342

    857 AACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGT 906
          |||   :::|||||||||||||||
    342 roLeu...ValGlyGluLeuValAsp........................ 349

    907 TCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACG 956
                           |||:::   :::|||:::::::::::::
    350 ..................IleGlyPheAlaAlaTyrAsnPheValGluSe 360

    957 CTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGT 1006
              |||   :::::::|||:::::::::::::   ::::::      :
    360 rIleIleAsnLeuPheGlnValValHisAsnSerTyrAsnArgProAlaT 377

   1007 TCGGCTGGGTTGCGCCG.............CAGCCG.........GACAAA 1035
        ::          :::|||             ||||||          |||
    377 yr.........SerProGlyHisLysThrGlnProPheLeuHisAspGly 390

   1036 TACTCCATCACGCGCACCACTCTC......................... 1059
        |||:::::::::   :::|||:::
    391 TyrAlaValSerTrpAsnThrValGluAspSerIleIleArgThrGlyPh 407

   1060 .............GGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGA 1096
                      ||||||   :::|||                   |
    407 eGlnGlyGluSerGlyHisAspIleLys.................IleT 418

   1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
        |||||   |||            :::   :::|||||||:::
    418 hrAlaGluAsn.............ThrProLeuProIleAlaGlyVal... 429

   1147 CGCGTAATGCCGTTGGACATCCTGCCTACCTTG 1179
                           :::||||||||||:::
    430 ..................LeuLeuProThrIle 434
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21970

seq_documentation_block:

ID ABG21970 standard; Protein; 575 AA.
XX
AC ABG21970;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21961.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86157.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52329; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 575 AA;

alignment_scores:

```
           Quality:  105.00              Length:    519
             Ratio:  0.515                 Gaps:     24
 Percent Similarity:  39.306   Percent Identity: 19.653
```

alignment_block:

US-09-303-518D-131 x ABG21970 ..

Align seg 1/1 to: ABG21970 from: 1 to: 575

```
     31 CCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTAC 80
        |||   :::|||::::         |||        |||||||||
     46 ProGlySerGlyLysSerProProValAlaSerGlyGlyProAla..... 60

     81 CGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGA 130
                                                 :::|||
     61 ..............................ArgAlaLysProHisArgA 67

    131 AAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180
          :::      :::   |||   :::   |||
     67 laValHisValSerProAlaProSerGlyGly.................. 77
```

```
181 AAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGC 230
        :::||||||   |||         |||||||:::      :::
 78 ......SerProGlyAspValAlaGluLeuProAlaAlaValAlaValAr 92

231 CGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCG 280
     |||   :::   |||:::   :::   |||   :::   |||   :::
 92 gAla***GlnAlaGlySerGlyGlnArgLeuProAlaProValCysValG 109

281 TTGAA..............GGCAACGACGAAATCGAGTTCGAACGCTAC 315
    |||                 ::::::      :::   ::::::
109 lyGluMetGlyThrLeuCysProGlyGluProAlaAspValHisGlnVal 125

316 GTA..............CCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
    |||              |||:::|||      :::   ::::::   ::
126 ValCysGlyLeuSerGluProArgAlaProHisArgProGlyGlnProGl 142

351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
    ::::|||   :::   :::   ||||||    :::|||
142 nAlaArgAlaHisGlyValProSerGly.GlnSerLeu............ 154

401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCT.. 448
                        |||   :::|||:::   :::   |||
155 .................CysProAlaGlnProGly***AspAspSerAl 165

449 .......TCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTAC 491
         ||||||      |||:::                ||||||
165 aLysValSerSerTyrSerTrpSerGlnSerTrpSerSerProThrAlaS 182

492 GGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTAT 541
      ::::::|||::::::|||||   :::   :::|||:::|||
182 erThrAspSerSerArgProProHisSerLeuGlyAlaSerCysHis... 197

542 TGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTA.........AAGCA 582
            |||:::   :::         :::         :::
198 ............ProGlyHisGlnArgSerLeuMetProProSerArg** 210

583 GCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGA 632
     :::|||   :::         :::              |||
210 *ArgAlaThrSerThrAlaTrpArgAlaArgThrAlaSerThrProArgC 227

633 ATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCA 682
      |||:::   |||      ||||||      :::   :::   :::|||
227 ysLeuGlyProArg......ProAlaProTrpSerCysAlaSerLeuSer 241

683 TCGAGCCAGTCGGCGCGAATAAAACCGTGT...GGACCATCAATTATCAA 729
    ||||||:::      :::   :::|||      :::|||:::   :::::
242 SerSerHisArgProAsnThrGlnProSerProSerProThrSerValAr 258

730 GACGTGATTGCTATCGGACGTTTGTTCG..................... 757
    :         |||:::         |||
258 gArgAlaArgLeuAlaThrProSerSerHisGlyTrpAlaLeuAlaProT 275

758 .....TAACAGGCCGTCTGAATACCGAGCGCGTGGTTG............ 790
          :::|||       |||::::::|||
275 rpThrProMetAlaProThrGlnProThrThrTrpThrArgTrpGlyLys 291

791 ..................................CCTTGGGCGGCCTG 804
                                     ||||||::::::||
292 ThrValSerGlyArgGlnMetAsnThrTrpArgArgProTrpSerThrCy 308

805 C....................AAGTCAACAAACCGCGCCTCTTGCGTAC 833
    |                   |||   ::::::   :::|||       |
308 sAlaArgTyrSerGlySerAlaLysArgSerSerGlySerSerHisSerP 325

834 CGTTTTGGGTGCGAAGGTGT............CTCAACTTACCGCCGGCG 871
    ||:::         |||                 |||   |||   |||
325 roTrpAlaProProArgMetArgMetGluLysSerAsnSerProThrAla 341

872 AATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGT 921
    :::|||:::         ::::::      |||
342 SerTrpValArgThrAspSerSerLeuArgPro.............. 352

922 GCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGAT 971
```

```
                                  |||   :::|||
353 ..........................TrpGlyGlyThr..... 356

972 TTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGC 1021
                            ::::::   :::|||:::   ||||||:::
357 ................ArgSerSerMetGlyCysGlu...GlyLeuLys 366

1022 CGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATT..... 1066
       |||      :::|||   |||   :::|||   :::|||:::
367 LeuSerThrArgGlyThrArgSerCysSerProSerGlyAlaMetArgSe 383

1067 ....TCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGG 1112
         :::   ::::::::::|||   :::   |||   |||||||||:::
383 rProAlaTrpSerAlaHisSerLeuGlyCysArgLeuProSerThrThrA 400

1113 CGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGG 1162
                                                  ||||||
400 spLeuGlnAsp..............................ArgTrp 405

1163 ACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGAC 1212
           |||:::|||   :::   :::       :::|||:::      ||
406 ArgLeuCysValProGlyMetThrSerSerAlaAlaSerValAlaThrTh 422

1213 AGCGCGCAGGCTT..................................... 1225
     |::::::::::|||
422 rSerGlnAsnLeuGlyTrpProAlaGlyThrCys***AlaLeuTrpGlyG 439

1226 ........TGGGTTGCT.............................. 1234
             |||   |||
439 lyGlyArgTrpProAlaThrProAlaAlaGlnAlaGlnProAlaLeuPro 455

1235 ..................TGGAAT.........TGGACGAAGAAGACCTC 1257
                        |||:::         |||   ::::::
456 GlyGlnLeuProThrArgTrpSerArgCysCysTrpProSerSerTrpPr 472

1258 GCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTT 1307
      ||||||:::||||||            :::   :::   ||||||||||
472 oLeuCysSerAlaSer............GlyProSerHisAlaArgCysC 485

1308 GCGCA 1312
     ||:::
485 ysSer 486
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26240

seq_documentation_block:

```
ID   AAB26240 standard; Protein; 1081 AA.
XX
AC   AAB26240;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit NMDAR1A #18.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; ionotropic;
KW   glutamate receptor; drug screening; animal model; disease diagnosis;
KW   genetic screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 116
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 140
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 162
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 183
FT                   /label= unknown
FT                   /note= "encoded by TGA"
```

```
FT   Misc-difference 258
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 274
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 300
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 313
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 337..338
FT                   /label= unknown
FT                   /note= "encoded by TGATGA"
FT   Misc-difference 431
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 561
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 593..594
FT                   /label= unknown
FT                   /note= "encoded by TGATGA"
FT   Misc-difference 692
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 697
FT                   /label= unknown
FT                   /note= "encoded by TAA"
FT   Misc-difference 858
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 1001
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 1058
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 1076
FT                   /label= UNKNOWN
FT                   /note= "encoded by TGA"
FT                   /note= "the coding sequence of this protein contains a
FT                   number of in-frame stop codons which are represented
FT                   by Xaa in the protein sequence"
XX
PN   US6111091-A.
XX
PD   29-AUG-2000.
XX
PF   29-SEP-1997;   97US-0940086.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Daggett LP,  Lu C;
XX
DR   WPI; 2000-578607/54.
DR   N-PSDB; AAA95033.
XX
PT   Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
PT   for identifying mutations and for developing drugs against various
PT   disease states  -
XX
PS   Example 3; column 633-672; 205pp; English.
XX
CC   The present sequence is a subunit (designated NMDAR1A) of the human
CC   N-methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate
CC   receptor which contains cation-specific ligand-gated ion channels. The
CC   protein and its coding sequence can be used in disease diagnosis and in
CC   research to identify other, similar proteins. They can also be used as
CC   probes, for example in genetic screening, and in drug screening, as well
CC   as enabling the production of animal disease models.
XX
SQ   Sequence   1081 AA;


alignment_scores:
           Quality:  105.00              Length:    483
             Ratio:  0.593                Gaps:     28
 Percent Similarity:  36.646   Percent Identity: 23.602

alignment_block:
 US-09-303-518D-131 x AAB26240   ..

 Align seg 1/1  to: AAB26240  from: 1  to: 1081

      26 ATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCC 75
         ::::::|||||||||         :::   :::            |||||
     705 LeuAlaProSerArgIleGlyValAlaAlaAlaValArgProHisArgPr 721

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC 125
         |       :::   |||   ||||||            ||||||   |||
     721 oProAlaArgProArgGlyLeuAlaPro...AlaHisAlaCysProPro. 736

     126 GATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGG...CCAAGTGCTGT 172
                        :::::::||| |||:::|||   |||   |||
     737 .............ProThrArgPro.GlnSerArgAlaProArgAlaGl 748

     173 TTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGC 222
             ||||||         :::|||   :::   |||||||||:::|||
     748 yAspArgGlnThrGlyValAlaArgArgLeu...CysAlaGlyLeuArg. 763

     223 AAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGT...ACTTCAGTCAGT 269
         :::   |||   :::|||   |||:::::::|||         :::|||
     764 ..SerProArgAlaAlaProArgArgArgGlyArgProCysProThrSer 779

     270 CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTAC 319
             :::||||||       |||      :::|||
     780 ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyPr 796

     320 CTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
             :::||||||      ::::::      |||:::|||
     796 oGlyThrAlaGlyGlyThrSerArgProProSerGlyPro......... 809

     370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCC 419

     809 ..................................................... 809

     420 TGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCA 469

     809 ..................................................... 809

     470 ATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC 519
                 |||||||||                              |||   :
     810 .........CysArgPro.....................ArgAlaV 815

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         ::   |||   |||      :::||||||   |||   :::      |||
     815 alThrThrAlaProPhe...LeuGluProThrAspProAla...AlaPro 829

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGC...TGCCA 616
                 |||||||||      |||:::         |||   |||::
     830 SerSerArgSerSerArgSerProArgSerTrpArgThrCysArgCysSe 846

     617 ATATCGAAACACATGAATTTGG.........CGGCCCGCATCC....... 650
         :   |||::::::      |||         ||||||::::::
     846 rValArgSerSer......TrpProGlyGlyArgProCys***ThrArgP 861

     651 ................................TGCCGGCTTGAGTGG 665
                                        |||       |||
     861 roGlyProGlyAlaArgAlaArgValThrLeuProCysProAlaProTrp 877

     666 CACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGA 715
```

```
 877 ..................................................  877

 716 CCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGC  765
                                                  ||||
 878 ..........................................ProArgP  880

 766 CGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAA  815
     |||||         ||||||      ||||||   ||||||
 880 roSerLeuGlyProAlaArgCysProLeuGlyAlaProAlaProProAla  896

 816 ACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCG.865
       ||||||   |||   |||                :::::::::::||
 897 ProAlaProThrAlaThrArg............ProAlaGlyAlaTrpAr 909

 866 CCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTG  915
     |||||||                          :
 909 gArgArg.....................................S  912

 916 AACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAA  965
     ::|||||   |||
 912 erArgCysAlaCys.................................. 916

 966 TCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGG 1015
           |||         ||||||   :::||||||   |||   ||
 917 .........ArgSerThrGlyArgProAlaArgArgAlaSerArgGlnGl 930

1016 T..................................................  1016
     |
 930 yProProProGlySerThrAspSerThrSerAlaCysThrProThrProT 947

1017 ..TGCGCCGCAGCCGGACAAATACTCC.................ATCAC 1046
       |||   ::::||||||   ::: |||                :::||
 947 hrCysHisCysAlaGlyGlyLeu.SerValLeuThrPheHisProValTh 963

1047 GCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGA 1096
     |   |||:::   |||           :::   |||:::      ::::
 963 rAlaThrAlaProGlySerProAlaProGlyGlyLeuTrpGlyThrAlaA 980

1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
     ::::::::   |||      :::|||         :::|||   :::
 980 laGlyLeuTrpGlyTrpAlaGlnAlaThrGluThrValGlyAspTrpThr 996

1147 CGC......GTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGA 1190
     |||      |||   |||:::          :::         |||||
 997 ArgSerAlaVal***ProValGlyArgLysAlaSer.........ArgAs 1010

1191 TTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAAT 1240
     |         ||||||                    ||| :::
1010 pProAlaProGlyAsp.....................Gly.SerProVal 1019

1241 TGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATAC 1290
     |||:::::::|||   :::|||   |||:::::::::: |||:::|||:::
1020 TrpSerGlnLysCysGluLeuSerAlaThrGlnAla.ProSerGlnLeuA 1036

1291 GAATACGGCCCGCT.........GTTGCGCAAAGTGCTGGA 1322
           ||||||         |||   ::::::||||||
1036 spSerLeuProAlaThrValArgValLysArgGlnAlaGly 1049
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239

seq_documentation_block:

```
ID   AAB26239 standard; Protein; 1232 AA.
XX
AC   AAB26239;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit NMDAR1A #17.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; ionotropic;
KW   glutamate receptor; drug screening; animal model; disease diagnosis;
KW   genetic screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 280
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 304
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 326
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 347
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 422
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 438
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 464
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 477
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 501..502
FT                   /label= unknown
FT                   /note= "encoded by TGATGA"
FT   Misc-difference 590
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 712
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 744..745
FT                   /label= unknown
FT                   /note= "encoded by TGATGA"
FT   Misc-difference 843
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 848
FT                   /label= unknown
FT                   /note= "encoded by TAA"
FT   Misc-difference 1009
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 1152
FT                   /label= unknown
FT                   /note= "encoded by TAG"
FT   Misc-difference 1209
FT                   /label= unknown
FT                   /note= "encoded by TGA"
FT   Misc-difference 1227
FT                   /label= UNKNOWN
FT                   /note= "encoded by TGA"
FT                   /note= "the coding sequence of this protein contains a
FT                   number of in-frame stop codons which are represented
FT                   by Xaa in the protein sequence"
XX
PN   US6111091-A.
XX
PD   29-AUG-2000.
XX
PF   29-SEP-1997;   97US-0940086.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (MERI ) MERCK & CO INC.
XX
```

```
PI   Daggett LP,  Lu C;
XX
DR   WPI; 2000-578607/54.
DR   N-PSDB; AAA95032.
XX
PT   Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
PT   for identifying mutations and for developing drugs against various
PT   disease states  -
XX
PS   Example 3; column 255-264; 205pp; English.
XX
CC   The present sequence is a subunit (designated NMDAR1A) of the human
CC   N-methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate
CC   receptor which contains cation-specific ligand-gated ion channels. The
CC   protein and its coding sequence can be used in disease diagnosis and in
CC   research to identify other, similar proteins. They can also be used as
CC   probes, for example in genetic screening, and in drug screening, as well
CC   as enabling the production of animal disease models.
XX
SQ   Sequence   1232 AA;


alignment_scores:
            Quality:  105.00              Length:    483
              Ratio:  0.593                Gaps:     28
 Percent Similarity:  36.646   Percent Identity: 23.602

alignment_block:
 US-09-303-518D-131 x AAB26239   ..

 Align seg 1/1  to: AAB26239  from: 1  to: 1232

      26 ATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCC 75
         ::::::|||||||||       :::   :::          |||||
     856 LeuAlaProSerArgIleGlyValAlaAlaAlaValArgProHisArgPr 872

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC 125
         }      :::   |||   ||||||           ||||||   |||
     872 oProAlaArgProArgGlyLeuAlaPro...AlaHisAlaCysProPro. 887

     126 GATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGG...CCAAGTGCTGT 172
                       :::::::||| |||:::|||   |||   |||
     888 .............ProThrArgPro.GlnSerArgAlaProArgAlaGl`899

     173 TTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGC 222
         ||||||         :::|||   :::   |||||||||:::|||
     899 yAspArgGlnThrGlyValAlaArgArgLeu...CysAlaGlyLeuArg. 914

     223 AAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGT...ACTTCAGTCAGT 269
           :::   |||   :::|||   |||:::::::|||         :::|||
     915 ..SerProArgAlaAlaProArgArgArgGlyArgProCysProThrSer 930

     270 CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTAC 319
           :::||||||      |||      :::|||
     931 ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyPr 947

     320 CTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATT 369
           :::||||||      ::::::      |||:::|||
     947 oGlyThrAlaGlyGlyThrSerArgProProSerGlyPro......... 960

     370 CAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCC 419

     960 ..................................................  960

     420 TGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCA 469

     960 ..................................................  960

     470 ATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGAC 519
                 |||||||||                          |||   :
     961 .........CysArgPro.....................ArgAlaV 966

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569

         ::   |||   |||      :::||||||   |||   :::      |||
     966 alThrThrAlaProPhe...LeuGluProThrAspProAla...AlaPro 980

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGC...TGCCA 616
                 |||||||||      |||:::          |||   |||::
     981 SerSerArgSerSerArgSerProArgSerTrpArgThrCysArgCysSe 997

     617 ATATCGAAACACATGAATTTGG.........CGGCCCGCATCC....... 650
         :   |||::::::      |||         ||||||::::::
     997 rValArgSerSer......TrpProGlyGlyArgProCys***ThrArgP 1012

     651 ................................TGCCGGCTTGAGTGG 665
                                         |||         |||
    1012 roGlyProGlyAlaArgAlaArgValThrLeuProCysProAlaProTrp 1028

     666 CACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGA 715

    1028 ..................................................  1028

     716 CCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGC 765
                                                       ||||
    1029 ....................................ProArgP 1031

     766 CGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAA 815
         |||||         ||||||      ||||||   ||||||
    1031 roSerLeuGlyProAlaArgCysProLeuGlyAlaProAlaProProAla 1047

     816 ACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCG 865
           ||||||   |||   |||              ||||||||||||
    1048 ProAlaProThrAlaThrArg.....ProAlaGlyAlaTrpAr 1060

     866 CCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTG 915
         |||||||                                      :
    1060 gArgArg.....................................S 1063

     916 AACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAA 965
         ::||||||   |||
    1063 erArgCysAlaCys.............................. 1067

     966 TCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGG 1015
               |||         ||||||   :::||||||   |||   ||
    1068 .........ArgSerThrGlyArgProAlaArgArgAlaSerArgGlnGl 1081

    1016 T.................................................  1016
         |
    1081 yProProProGlySerThrAspSerThrSerAlaCysThrProThrProT 1098

    1017 ..TGCGCCGCAGCCGGACAAATACTCC.................ATCAC 1046
           |||   :::||||||   ::: |||                :::||
    1098 hrCysHisCysAlaGlyGlyLeu.SerValLeuThrPheHisProValTh 1114

    1047 GCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGA 1096
         |   |||:::   |||            :::   |||:::      ::::
    1114 rAlaThrAlaProGlySerProAlaProGlyGlyLeuTrpGlyThrAlaA 1131

    1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAG 1146
         ::::::::   |||       :::|||         :::|||   :::
    1131 laGlyLeuTrpGlyTrpAlaGlnAlaThrGluThrValGlyAspTrpThr 1147

    1147 CGC......GTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGA 1190
         |||      |||   |||:::          :::         |||||
    1148 ArgSerAlaVal***ProValGlyArgLysAlaSer.........ArgAs 1161

    1191 TTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAAT 1240
         |         ||||||                        ||| :::
    1161 pProAlaProGlyAsp.....................Gly.SerProVal 1170

    1241 TGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATAC 1290
         |||::::::|||   :::|||   |||:::::::: |||:::|||:::
    1171 TrpSerGlnLysCysGluLeuSerAlaThrGlnAla.ProSerGlnLeuA 1187

    1291 GAATACGGCCCGCT.........GTTGCGCAAAGTGCTGGA 1322
                |||||||         |||   ::::::||||||
```

1187 spSerLeuProAlaThrValArgValLysArgGlnAlaGly 1200